



**Evaluación de modelos de predicción de la demanda de exportación de café
verde en Colombia: un enfoque desde la inteligencia de negocios y cadenas de
suministro**

Jennifer Agámez Hernández

Katherine Del Pilar Padilla Santos

Sandra Janeth Beltrán Aguiar

Universidad Ean

Facultad de ingeniería

Maestría en Inteligencia de Negocios

Maestría Gerencia de la Cadena de Abastecimiento

Bogotá, Colombia

Enero 2024

Evaluación de modelos de predicción de la demanda de exportación de café verde en Colombia: un enfoque desde la inteligencia de negocios y cadenas de suministro

Jennifer Agámez Hernández

Katherine Del Pilar Padilla Santos

Sandra Janeth Beltrán Aguiar

Trabajo de grado presentado como requisito para optar al título de:

Magister en Inteligencia de Negocios y Magister en Gerencia de la Cadena de Abastecimiento

Director (a):

Carolina María Luque Zabala

Modalidad:

Monografía

Universidad Ean

Facultad de Ingeniería

Maestría en Inteligencia de Negocios

Maestría en Cadena de Abastecimiento

Bogotá, Colombia

Enero 2024

Fuente de aceptación:

Firma del jurado

Firma del jurado

Firma del director del trabajo de grado

Ciudad, día/mes/año

Dedicatoria

Jennifer.

Quiero dedicar este trabajo a las personas que me acompañan tras bambalinas, a mi esposo Brayan por su constante apoyo, a mi mamá Mirta y mi papá Elder en el cielo; mis hermanas María Fernanda y Luz Teresa, a mi incondicional Kiki y a Dios. Por último y no menos importante, quiero dedicarlo a mí por dar siempre lo mejor de mí y por creer en mí.

Katherine.

Este Trabajo lo dedico a Fabian, mi esposo y a mis hijos Tomas, Valentina y Mariana, quienes me dieron todo su apoyo y su paciencia para hoy poder alcanzar esta meta. A mi familia por siempre confiar en mi e impulsarme a ser mejor cada día.

Sandra.

El apoyo de mi esposo Diego y mis hijas Sara y Victoria fue fundamental para alcanzar este logro, por eso quiero dedicar este trabajo a ellos, a Dios, a mis padres y a mi hermana Paola, por enseñarme el compromiso y la perseverancia, a hacer posible lo que parece imposible.

Agradecimientos

Queremos agradecer a cada una de nuestras familias, madres, padres, hermanos, esposos e hijos que estuvieron dándonos apoyo, sin ustedes no es posible este logro.

Extendemos nuestro agradecimiento a nuestra directora y mentora de tesis Carolina Luque por apoyo y direccionamiento en todo el proceso de esta monografía. Su guía estricta pero acertada, su paciencia y su motivación han sido fundamentales para nuestra formación como magísteres y nos lleva en un proceso de mejora continua cada día.

También agradecimientos a la Universidad EAN por fomentar y brindarnos una educación de calidad.

Resumen

El café, un producto básico comercializado en todo el mundo, desempeña un papel vital en la economía de Colombia como uno de los principales productores y exportadores. Esta monografía, tiene como propósito evaluar modelos de predicción de demanda de exportación de café verde en Colombia para apoyar la toma de decisiones en la cadena de suministro del producto, desde un enfoque de inteligencia de negocios. Se utilizaron datos de producción, exportación, precios internacionales y tasa de cambio para construir y evaluar cuatro modelos de aprendizaje automático.

Los resultados muestran que el modelo de Máquinas de Soporte Vectorial presenta el mejor rendimiento para la predicción de las exportaciones de café verde en Colombia. Estos resultados ponen en relieve el potencial de los modelos de aprendizaje automático dentro de un marco de inteligencia de negocios para mejorar la toma de decisiones en la cadena de suministro del café. La inclusión de modelos predictivos permite a las partes interesadas adaptar las estrategias a diversos escenarios, fomentando las respuestas proactivas a los cambios en la demanda y las condiciones del mercado.

Palabras clave: Café verde, demanda de exportación, Inteligencia de negocios, aprendizaje automático, cadena de suministro.

Abstract

Coffee, a commodity traded worldwide, plays a vital role in Colombia's economy as a major producer and exporter. The purpose of this monograph is to evaluate predictive models of export demand for green coffee in Colombia to support decision-making in the supply chain of the product, from a business intelligence approach. Production, export, international prices, and exchange rate data were used to build and evaluate four machine learning models.

The results show that the Support Vector Machines model presents the best performance for the prediction of green coffee exports in Colombia. These results highlight the potential of machine learning models within a business intelligence framework to improve decision-making in the coffee supply chain. The inclusion of predictive models allows stakeholders to adapt strategies to various scenarios, encouraging proactive responses to changes in demand and market conditions.

Keywords: Green coffee, Export demand, Business Intelligence, Machine Learning, Supply Chain, Forecasting.

Contenido

1	Introducción	12
2	Objetivos	14
2.1	<i>Objetivo General</i>	14
2.2	<i>Objetivos específicos</i>	14
3	Justificación	15
4	Marco Teórico	17
4.1	<i>La industria del café en Colombia</i>	17
4.2	<i>Cadena de Suministro</i>	19
4.2.1	Cadena de suministro del café en Colombia.....	21
4.2.2	Factores económicos que influyen en la demanda de exportación del café verde	24
4.2.3	Sostenibilidad en la cadena del suministro del café.....	27
4.3	<i>Inteligencia de negocios</i>	28
4.3.1	Modelo de inteligencia de negocios	29
4.3.2	Implementaciones de inteligencia de negocios en el sector del cafetero	32
4.4	<i>Modelos de predicción</i>	37
4.4.1	Modelos de aprendizaje automático	38
4.4.2	Selección de variables.....	39
4.4.3	Datos de entrenamiento y prueba.....	40
4.4.4	Sobreajuste y subajuste.....	41

EVALUACIÓN DE MODELOS DE PREDICCIÓN DE LA DEMANDA DE EXPORTACIÓN DE CAFÉ VERDE EN COLOMBIA: UN ENFOQUE DESDE LA INTELIGENCIA DE NEGOCIOS Y CADENA DE SUMINISTRO	14
4.4.5 Algoritmos utilizados para la predicción de la demanda.....	41
4.4.6 Métricas para comparar modelos y evaluar su rendimiento	45
5 Hipótesis	46
6 Fuentes de datos y variables	47
6.1 <i>Fuentes de datos</i>	47
6.2 <i>Variables</i>	48
7 Diseño metodológico	50
7.1 <i>Fase 1</i>	50
7.2 <i>Fase 2</i>	51
7.3 <i>Fase 3</i>	51
7.4 <i>Fase 4</i>	52
8 Trabajo de Campo	52
8.1 <i>Proceso ETL, almacenamiento y modelo de datos</i>	52
8.2 <i>Caracterización de factores económicos que influyen en el comportamiento de la demanda de exportación de café verde en Colombia</i>	54
8.3 <i>Preprocesamiento de los datos</i>	61
8.4 <i>Construcción del modelo de ML</i>	62
8.4.1 <i>Modelo de regresión lineal múltiple</i>	63
8.4.2 <i>Regresión de Ridge</i>	63
8.4.3 <i>Modelo de Red elástica (Elastic Net)</i>	63

EVALUACIÓN DE MODELOS DE PREDICCIÓN DE LA DEMANDA DE EXPORTACIÓN DE CAFÉ VERDE EN COLOMBIA: UN ENFOQUE DESDE LA INTELIGENCIA DE NEGOCIOS Y CADENA DE SUMINISTRO	15
8.4.4 Modelo Máquinas de Soporte Vectorial (SVM)	64
9 Discusión de resultados	65
10 Orientaciones metodológicas para la predicción de la demanda de exportación de café verde en Colombia bajo el enfoque de BI.....	71
11 Conclusiones y trabajos futuros.....	75
11.1 Conclusiones.....	75
11.2 Trabajos futuros	77
12 Referencias.....	79
13 Anexos	88

Lista de Figuras

Figura 1 Departamentos donde se cultiva café en Colombia	18
Figura 2 Esquema de una cadena de abastecimiento.....	20
Figura 3 Proceso de poscosecha del café	22
Figura 4 Café verde empacado al vacío.....	23
Figura 5 Cadena de suministro del café en Colombia	24
Figura 6 Arquitectura de capas propuesto	30
Figura 7 Ecuación de búsqueda Web of Science.....	33
Figura 8 Etapas para la implementación de un modelo de Machine learning	38
Figura 9 Algoritmos de ML seleccionados.....	42
Figura 10 Hiperplano de separación óptimo construido con SVM	44
Figura 11 Volumen de exportaciones colombianas de café	48
Figura 12 Visualización de primeras filas de la base de datos construida	53
Figura 13 Construcción del modelo de datos	53
Figura 14 Análisis de la producción, exportaciones e inventarios de café verde	55
Figura 15 Análisis de precios internacionales del café colombiano	56
Figura 16 Diagrama de cajas de variables	57
Figura 17 Histograma de variables analizadas.....	58
Figura 18 Correlación de variables	60
Figura 19 Resultados de modelos.....	65
Figura 20 Comparación de los resultados.....	66
Figura 21 Histograma de residuales de modelos	67
Figura 22 Correlación de predicciones vs. Valores reales de los modelos	68
Figura 23 Proceso de toma de decisiones para la cadena de suministro del café verde en Colombia	69

Figura 24 Ejemplo de CMI con integrando el modelo de la predicción de la demanda.

..... 70

Figura 25 Arquitectura de BI – propuesta desde las orientaciones metodológicas..... 73

Lista de Tablas

Tabla 1 Factores Económicos que afectan los precios de referencia del café.....	26
Tabla 2 Resumen de revisión bibliográfica.....	33
Tabla 3 Resumen de variables	49
Tabla 4 Resumen descriptivo de variables.....	57

1 Introducción

El café es uno de los productos básicos de mayor comercialización en el mundo (División de Estadística de la ONU, 2022). Colombia, entre los principales productores de este grano en el mundo, exportó un total de 11.408.241 sacos de café de 60 kg en 2022, a destinos como Estados Unidos, Canadá y Alemania (*Estadísticas cafeteras - Federación Nacional de Cafeteros de Colombia, s/f*). Se destaca que un 83 % del total de sacos exportados desde Colombia corresponde a café verde, lo que representó para ese mismo año, un valor de 3.478 millones de dólares (*Estadísticas cafeteras - Federación Nacional de Cafeteros de Colombia, s/f*).

Estos volúmenes de exportación del café están influenciados por factores económicos externos como la cotización internacional del grano, tal como lo señala Acevedo et al., (2020) y la tasa representativa del mercado, y, como el eje central de este trabajo, el comportamiento de la oferta y la demanda (Jiménez y Soto, 2018). Estas variables son fluctuantes creando un entorno incierto para productores y exportadores, exponiéndolos a riesgos, lo que subraya la necesidad de herramientas que proporcionen un mayor entendimiento del mercado, su dinámica actual y predicciones de su comportamiento futuro (Wang et al., 2021).

La predicción de la demanda no solo responde a las necesidades de un mercado altamente volátil, también considera las condiciones logísticas requeridas para satisfacer esta demanda, en su mayoría internacional (Jiménez y Soto, 2018). En ese sentido, es vital caracterizar la cadena de suministro del café verde en Colombia, desde la producción hasta la exportación, proporcionando herramientas y conocimiento a los productores y exportadores, para facilitar la toma de decisiones (López et al., 2021).

Dada la relevancia del tema, la bibliografía destaca la importancia de examinar con mayor precisión el comportamiento de la demanda del café (Bacci et al., 2019; Kittichotsatsawat

et al., 2022). Aunque, se ha observado un incremento en la producción científica relacionada con modelos de predicción basados en inteligencia de negocios (Acevedo et al., 2020; Benos et al., 2021; Jiménez y Soto, 2018; Montoya et al., 2009), la revisión de la literatura contemporánea ha señalado que aún son escasos los estudios relacionados con la predicción de demanda de café verde en Colombia.

Con el objetivo de contribuir a este campo de estudio, este trabajo busca responder a la pregunta: ¿incluir modelos de predicción de demanda de exportación en una arquitectura de inteligencia de negocios contribuye a la toma de decisiones informadas por parte de los actores que participan en la cadena de suministro de café verde en Colombia? Para ello, se propone construir y evaluar modelos de predicción de demanda de exportación de café verde en Colombia para apoyar la toma de decisiones en la cadena de suministro del producto, desde un enfoque de la inteligencia de negocios (BI, por sus siglas en inglés), utilizando algoritmos de aprendizaje automático (machine learning, ML por sus siglas en inglés), a partir de datos relacionados con los volúmenes de producción, exportación, precios de referencia interna e internacional del café verde colombiano provenientes de la Federación Nacional de Cafeteros (FNC).

La estructura del documento inicia con la presentación de objetivos, justificación y un marco teórico centrado en la evaluación de modelos predictivos para la cadena de suministro del café verde en Colombia. A continuación, se plantea la hipótesis, las variables y las fuentes de información, así como la metodología utilizada en la investigación. Posteriormente, se presentan los resultados obtenidos con su respectivo análisis, seguidos por orientaciones metodológicas para la predicción de la demanda de café verde en Colombia, subsecuentemente se presentan las conclusiones y trabajos futuros.

2 Objetivos

2.1 Objetivo General

Evaluar modelos de predicción de demanda de exportación de café verde en Colombia para apoyar la toma de decisiones en la cadena de suministro del producto, desde un enfoque de inteligencia de negocios.

2.2 Objetivos específicos

- Elaborar una revisión de la literatura que guíe la construcción del modelo de predicción de demanda de exportación de café verde en Colombia desde el enfoque de la inteligencia de negocios y cadenas de suministro.
- Caracterizar los factores económicos que influyen en el comportamiento de la demanda de exportación de café verde en Colombia.
- Construir y evaluar modelos de predicción de la demanda de exportación de café verde en Colombia, utilizando técnicas de aprendizaje automático.
- Proponer las orientaciones metodológicas para la predicción de la demanda de exportación de café verde en Colombia bajo el enfoque de inteligencia de negocios.

3 Justificación

Colombia durante el período 2021 - 2022 se posicionó como el tercer productor de café en el mundo (United States Department of Agriculture, 2023b). De acuerdo con la Federación Nacional de Cafeteros de Colombia, en este período, Colombia produjo 23.6 millones de sacos de café de 60 kg, lo que representó el 7% del total de la producción mundial de este grano (Estadísticas cafeteras - Federación Nacional de Cafeteros de Colombia, 2023).

En cuanto a las exportaciones, Colombia es de los primeros tres países exportadores a nivel mundial, con 23.8 millones de sacos de 60 kg en el período 2021-2022, un 8% del total de café exportado mundialmente. Asimismo, siendo un líder productor y exportador, Colombia destaca como un significativo consumidor, con un consumo interno aproximadamente de 4.2 millones de sacos de café de 60 kg en ese mismo período (USDA, 2023).

Aunque Colombia es un gran productor de café con una fuerte presencia en el mercado internacional, durante los años 2021-2022, se vio obligado a importar 4.6 millones de sacos de café, principalmente de Brasil y Perú (USDA, 2023a). Esta situación resalta retos en la cadena agroindustrial, como la insuficiencia en la producción e inventarios, destacando la importancia de prepararse para las fluctuaciones del mercado (Kittichotsatsawat et al., 2022).

Ante estos desafíos, la gestión efectiva de la cadena de suministro es esencial para equilibrar el suministro y evitar excedentes o escasez en la producción de café. En este contexto, la previsión de la demanda de exportación se destaca como un elemento esencial para planificar de manera efectiva y sostenible la capacidad productiva futura (Wang et al., 2021). Esta anticipación se convierte en un pilar fundamental para los actores de la cadena del café, especialmente los exportadores, al permitir una planificación estratégica y una gestión precisa de la oferta en mercados internacionales, al mismo tiempo que garantiza el abastecimiento interno.

Con el objetivo de abordar estos desafíos, el presente trabajo propone la construcción y evaluación de modelos de predicción de la demanda externa de café verde en Colombia. Este enfoque integra herramientas de inteligencia de negocios (BI) y aprendizaje automático, proporcionando no solo una gestión más efectiva de las exportaciones dentro de la cadena de suministro del café, sino también un marco estratégico para adaptarse proactivamente a los cambios del mercado. Desde el punto de vista tecnológico, la aplicación conjunta de la inteligencia de negocios y las técnicas de machine learning ofrecen una visión global y estratégica, permitiendo la recopilación y evaluación precisa de datos en un mercado internacional dinámico (Joyanes, 2019). Asimismo, se espera que este enfoque facilite la toma de decisiones estratégicas, una utilización más eficiente de los recursos y la mejora en la precisión de los pronósticos de demanda externa del café verde colombiano (Kittichotsatsawat et al., 2021).

La presente monografía se enmarca en los temas de investigación del grupo ONTARE, bajo las líneas de investigación de los programas de Maestría en Cadena de Abastecimiento y Maestría en Inteligencia de Negocios de la Facultad de Ingeniería de la Universidad EAN.

4 Marco Teórico

4.1 La industria del café en Colombia

Se estima que existen 131 especies de café, de las cuales dos son las más ampliamente cultivadas: *Coffea Arábica* (comercialmente conocida como arábica) y *Coffea Robusta* (comercialmente conocida como robusta) (*History of Robusta*, 2013).

La especie arábica es la más cultivada en el mundo, representando el 57.5% del total de la producción mundial del grano y el 42.5% corresponde a robusta, según cifras del año 2022 de la ICO (Organización Internacional del Café, según nombre en inglés) (*Coffee Report and Outlook*, 2023), esta especie se caracteriza por producir bebidas suaves, con un aroma y acidez marcados, pero con sabor amargo moderado, por lo que se conoce como café suave. El café variedad *Robusta* es originario de África occidental, y es cultivado principalmente en África y Brasil, siendo este último, uno de los principales países productores de esta especie.

En tanto en Colombia es cultivada principalmente la especie arábica y es la base para que el país sea reconocido como uno de los principales productores de café suave en el mundo con un 13.9% de la producción mundial (*Coffee Report and Outlook*, 2023).

Desde principios del siglo XIX han sido miles de familias colombianas las que han tenido su sustento económico en torno a la cadena productiva del café, forjando las bases de la importancia de la cultura cafetera en Colombia (Martínez Chimbi, 2019). Según cifras de la Federación Nacional de Cafeteros (FNC, por sus siglas), existen unas 540 mil familias caficultoras, donde la gran mayoría se clasifican como pequeños caficultores cuyos cultivos, en promedio, no superan las 2 hectáreas (Regiones Cafeteras - Café de Colombia, 2020). Estos caficultores constituyen entonces, la base del capital social del sector cafetero, haciendo parte fundamental del tejido económico, social y cultural de Colombia.

4.2 Cadena de Suministro

La cadena de suministro (*Supply Chain*, SC, por sus siglas en inglés) se define como una estructura en la que se busca facilitar el movimiento de bienes entre los actores, organizaciones o individuos que la conforman (Chopra & Meindl, 2013; Velasco et al., 2023). Está diseñada de acuerdo con las características del proceso productivo con el fin de generar valor, manteniendo el monitoreo constante de las actividades que la integran y un uso adecuado de los recursos (Velasco et al., 2023). Para el cumplimiento de este propósito se requiere la participación de diferentes actores, que deben interactuar en cada una de las etapas una o varias veces, entre estos encontramos a los proveedores, fabricantes o productores, distribuidores, minoristas y consumidores.

El primero de los procesos que integra la cadena de suministro es la planeación; en este se desarrollan actividades estratégicas, de diseño, monitoreo y mejora continua. En segundo lugar, se encuentra el proceso de compra de materiales, materias primas y servicios que son necesarios para las siguientes etapas. El tercer eslabón es la producción, en esta etapa se realizan actividades de transformación de materiales, producción que debe ser coherente con la planeación establecida para satisfacer la demanda de los clientes. En cuarto lugar, se encuentra la etapa de distribución, en la que se debe administrar el almacenamiento, custodia y entrega a los minoristas o a los consumidores finales. Por último, está el proceso de devolución, que se puede dar a partir de los clientes (consumidores) o en otra etapa. Es importante mencionar que se han descrito los procesos como una secuencia, sin embargo, pueden suceder interacciones en diferentes niveles, de acuerdo con las necesidades de los participantes de la cadena (Mejía, 2023).

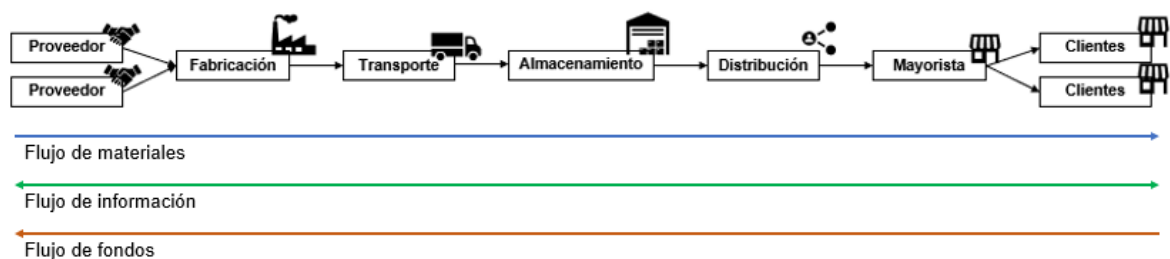
La coordinación de todas las etapas de la cadena de suministro, operando sincronizada y sinérgicamente, contribuye al cumplimiento de la meta de manera eficaz, evitando la generación de desperdicios, indicativo de altos niveles de inventario o, por el contrario,

incumplimiento en las entregas de pedidos a los clientes por escasez de productos. En este contexto, se destaca la importancia que tiene una adecuada planeación de la demanda, evitando la materialización de riesgos como los sobrantes o faltantes (Mejía, 2023).

Para que la SC logre su objetivo, es fundamental identificar tres tipos de flujo que conectan a las diferentes etapas de la cadena: Flujo de materiales, flujo de información y flujo de fondos, estos pueden ocurrir en una o ambas direcciones y pueden ser gestionados por más de un actor de la SC (Chopra y Meindl, 2013). El flujo de materiales tiene su punto de origen en los proveedores, que proporcionan las materias primas que se transformarán en productos terminados. El flujo de información se da en ambas vías y ocurre cuando se generan datos o documentos que contienen información valiosa para el funcionamiento de las operaciones. Finalmente, se encuentra el flujo de fondos, que corresponde al capital financiero que circula entre los integrantes de la cadena de suministro y a las ganancias resultantes de la gestión realizada (Mejía, 2023).

En la Figura 2 se encuentran representados los eslabones que componen una cadena de abastecimiento, así como los diferentes flujos que las interrelacionan facilitando el movimiento, la transformación y la distribución de bienes entre los diferentes actores (Taha, 2017).

Figura 2 Esquema de una cadena de abastecimiento



Fuente. Adaptado de “Etapas de una cadena de suministro”, por (Chopra y Meindl, 2013).

La integración de la SC con tecnologías emergentes de la información y la comunicación (TIC), como lo menciona Velasco et al., (2023), hace que se atenúe la frontera imaginaria que

existe entre el mundo digital y el mundo físico. Evidenciándose que el elemento de la cadena de suministro más susceptible de mejora ha sido el flujo de información, permitiendo la creación de ventajas competitivas para las organizaciones (Mejía, 2023) y un mejor posicionamiento en el mercado (Porter, 2007). Es así como la predicción de la demanda es uno de los procesos que se ha visto impactado por la implementación de nuevas herramientas tecnológicas favoreciendo la toma de decisiones en la cadena de suministro, ya que permite el acceso a información integral para optimizar el uso de los recursos disponibles (Heizer et al., 2009; Hofmann y Rutschmann, 2018).

4.2.1 Cadena de suministro del café en Colombia

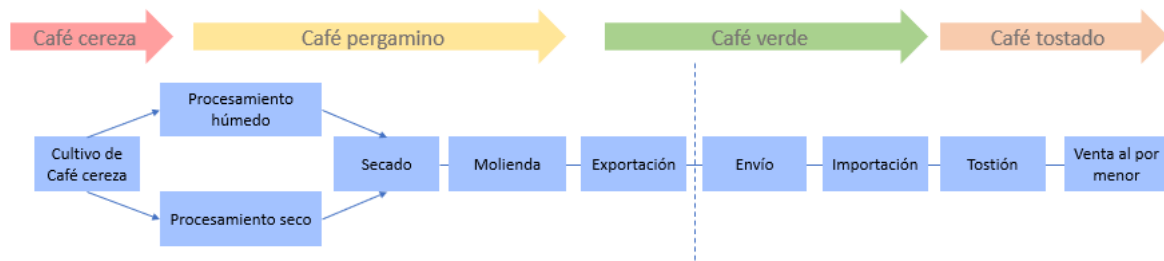
La cadena de suministro del café abarca todos los procesos y actividades que surgen desde que los productores adquieren la materia prima hasta la venta al consumidor. En el estudio adelantado por López et al., (2021) se encuentra caracterizada la cadena de suministro del café, en la que se identifica a los proveedores de agro insumos (semillas), productores en finca, las cooperativas de caficultores, las plantas de beneficio, las trilladoras, las tostadoras, exportadores, mayoristas y minoristas. En Colombia, la Federación Nacional de Cafeteros (FNC) juega un papel fundamental como representante nacional e internacional de café de Colombia, en su estructura cuenta también con Almacafé (Almacafé, 2018) como operador logístico y con el Centro de Investigaciones del Café (CENICAFÉ), creado con el fin de adelantar estudios relacionados con la producción, cosecha, beneficio, calidad, manejo y utilización de los subproductos derivados del café en Colombia (Centro Nacional de Investigaciones de Café, 2021) .

El primer eslabón de la cadena de suministro del café es la producción, que como se observa en la Figura 3, comienza con el cultivo y la recolección del fruto o cereza. El cual debe ser procesado en una planta de beneficio, en donde se procede con la remoción de la cáscara

y mucílago; etapa que puede ser realizada mediante remoción mecánica, o dejando que el café sea fermentado entre 12 a 18 horas dependiendo de variables como la temperatura del lugar, altura de la masa de café en el tanque de fermentación, el uso de agua, el grado de madurez y la cantidad de mucílago en el grano (*POSCOSECHA - Café de Colombia, 2020b*).

Subsecuentemente, se realiza el secado, el cual puede tardar hasta 10 días obteniendo así, el café pergamino seco, que puede ser vendido a dos actores, como son las cooperativas afiliadas a la FNC o a asociaciones y compradores privados (compradores locales, exportadores directos y multinacionales). La siguiente etapa es el proceso de trilla, que también se realiza en Colombia, en la cual se ejecutan además actividades de clasificación de calidad y de características organolépticas del café verde o almendra, el cual adquiere este nombre por la tonalidad que presenta luego de la trilla (ver Figura 4). Las pruebas realizadas para dicha selección se basan en lineamientos establecidos por la FNC y hacen parte de los procesos de Almacafé (López et al., 2021). En este punto, es importante mencionar, que solamente el café verde que cumple con los criterios de aceptación en la calidad es el que se aprueba para ser exportado, completando los procesos de secado y trillado (International Labour Organization, 2022). El café que no cumple con los criterios de calidad mencionados se aprueba para comercialización nacional y se conoce con pasilla.

Figura 3 Proceso de poscosecha del café



Fuente: elaboración propia, basado en Borrella et al., (2015)

Posteriormente y como última etapa, se encuentra la tostión, la cual puede llevarse a cabo tanto en Colombia como a nivel internacional. Es importante destacar que el 91% del café exportado es en su forma verde, por lo cual es tostado en máquinas situadas en los países que consumen la popular bebida (International Labour Organization, 2022).

Figura 4 Café verde empacado al vacío

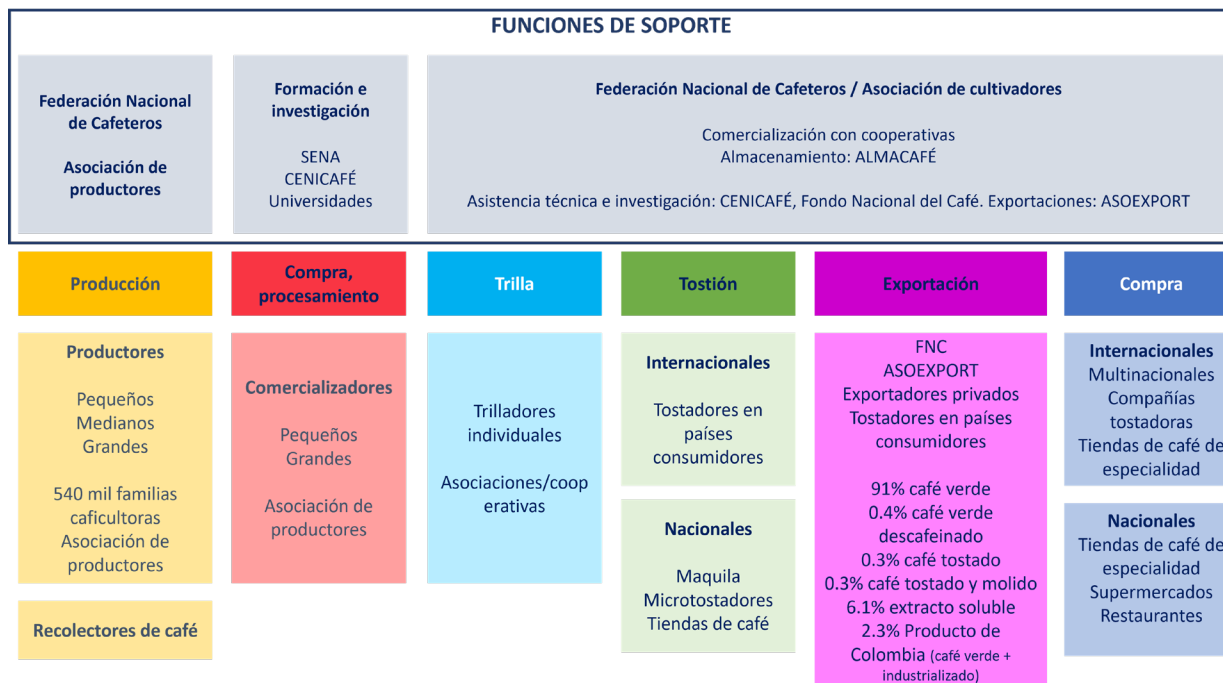


Fuente: Tomado de Federación Nacional de Cafeteros de Colombia (2022)

A lo largo de la cadena de abastecimiento del café, las grandes multinacionales del comercio ostentan actualmente un control predominante, un tema que suele generar considerables debates debido a sus vastos recursos y su acceso a financiamiento, que suele ser más económico. A pesar de ello, en numerosos casos, su involucramiento conduce a una mayor competencia en las fincas productoras, una dinámica que podría resultar en beneficios significativos para los caficultores (World Bank, 2016).

Del mismo modo, la FNC interviene como entidad privada y lo hace a través de políticas y programas sectoriales como el Fondo de Estabilización de Precios del Café Colombiano (FEPC), y como representante gremial ante el gobierno nacional, tal como se puede observar en la Figura 5. Su rol en la cadena abarca asistencia técnica a los caficultores, ayudándolos a mejorar sus prácticas de cultivo, producción y proporcionando un marco de protección al caficultor colombiano ante las variaciones en los precios de cosecha.

Figura 5 Cadena de suministro del café en Colombia



Fuente: Elaboración propia basado en International Labour Organization, (2022)

4.2.2 Factores económicos que influyen en la demanda de exportación del café verde

Con relación a la demanda de café, la literatura aborda esto considerando que los factores que influyen en su comportamiento son el precio del café, los ingresos de los consumidores y los bienes sustitutos disponibles (como el jugo de naranja, té y el cacao) (Ensayos sobre Economía Cafetera, 2014). Así mismo, se ha encontrado que factores como la cultura o hábitos de consumo y compra cumplen también un rol importante en dicha demanda (Ensayos sobre Economía Cafetera, 2014). No obstante, para efectos de análisis de la demanda externa en este trabajo, se expondrán los factores a saber: la producción, el precio internacional, la tasa de cambio, y el volumen de exportación. A continuación, se expone en detalle cada uno de estos.

4.2.2.1 Producción

La oferta del café (producción) y su calidad dependen de factores climáticos, inventario almacenado, los costos de la materia prima y factores socio-culturales asociados a los caficultores y a los canales de distribución, así como la estimulación a la innovación para ofrecer un mayor rendimiento del café verde que pueda ser exportado. De esta manera también se logra estimular la demanda, generando valor en la cadena e incrementando la preferencia del consumo del café colombiano en los consumidores de países importadores (Ensayos sobre Economías Cafeteras, 2014).

4.2.2.2 Precio internacional

Siendo el café un bien de consumo masivo, a menudo se enfrenta a las fluctuaciones de los precios internacionales, no obstante, es importante recalcar que, al no tratarse de un producto homogéneo, la Organización Internacional del Café (ICO, por sus siglas en inglés) propone cuatro tipos de precios:

- Físicos: corresponde al precio con el que se comercializa físicamente el café verde.
- Indicativos: se calculan para cada grupo de café. Colombia es tipo 1 (Arábicas suaves colombianos).
- Futuros: son las proyecciones de precios de café. Reflejan un estimado de disponibilidad y demanda futura del café. En el caso de los precios colombianos (Arábicas) las proyecciones se basan en la bolsa de Nueva York.
- Diferenciales: permite asociar los tipos de precios físicos y futuros.

En ese contexto, la OIC es la encargada de publicar de forma diaria el Precio Indicativo Compuesto (Grupo de Estudios Económicos-Superintendencia de Industria y Comercio, 2012). Además de los precios indicativos OIC, estos también están influenciados por el diferencial por calidad que es dado por la FNC, los cuales hacen parte de las fórmulas para calcular los precios de referencia (ver Tabla 1) (Ensayos sobre Economía Cafetera, 2014).

Tabla 1 Factores Económicos que afectan los precios de referencia del café

Factor	Unidad de medida	Fuente de información
Cotización del café en la bolsa de Nueva York	USD ¢/lb	Bolsa de Nueva York
Tasa de cambio representativa del mercado	\$/USD	Superintendencia Financiera de Colombia
Prima o diferencial por calidad	USD ¢/lb	Federación Nacional de Cafeteros de Colombia

Fuente: Elaboración propia.

La fijación de los precios del café (precio exdock, precio OIC y precio suave colombiano), al igual que en otros productos agroindustriales, depende de las variaciones entre la oferta y la demanda, es decir, cuando la oferta aumenta los precios internacionales tienden a bajar, para el caso contrario, cuando aumenta la demanda y hay escasez, los precios suben; esto dado la participación del café en el mercado internacional (Grupo de Estudios Económicos-SIC, 2012). Un mayor precio internacional del grano trae consigo un aumento en los precios de las materias primas, lo que incrementa el costo de la producción.

En respuesta de ello, los productores se ven obligados a tomar decisiones, algunos pequeños y medianos productores no pueden acceder a la materia prima y otros hacen uso de lo que logran adquirir, en ocasiones materiales poco eficientes, lo que afecta la calidad del café producido, y este, al no cumplir con las especificaciones requeridas, no logra llegar al mercado internacional (Ensayos sobre Economía Cafetera, 2014; Grupo de Estudios Económicos-SIC, 2012).

4.2.2.3 Tasa de cambio

En cuanto a la tasa de cambio, para Colombia la cantidad de pesos colombianos que se pagan por un dólar estadounidense es otro factor que impacta el precio internacional del café.

Cuando aumenta la tasa de cambio, se hacen más costosas las transacciones en el mercado, en la medida que se devalúan las divisas frente al dólar.

Así mismo, fenómenos como la revaluación del peso y el descenso de la producción también tienen su efecto en la disminución de ingresos por exportaciones, en consecuencia, los productores perciben cierta favorabilidad incrementando la producción, lo que genera un aumento en la oferta, y resultado de ello, disminución de los precios internacionales (Grupo de Estudios Económicos-SIC, 2012).

4.2.2.4 Exportaciones

Es posible identificar, a partir de las estadísticas de la FNC que en la medida que la producción aumenta, lo hacen las exportaciones (Estadísticas cafeteras - FNC, s/f). En dichas estadísticas, es posible apreciar que más del 80% del café verde producido es exportado a países como Estados Unidos (con cerca del 40% de exportaciones), seguido por la Unión Europea y Japón (USDA, 2023a). Este café es exportado en sacos de 60 kg, este comercio internacional del grano también se ve altamente influenciado por el ingreso de nuevos caficultores vinculados a programas de cafés especiales (Grupo de Estudios Económicos-SIC, 2012).

4.2.3 Sostenibilidad en la cadena del suministro del café

La sostenibilidad es un aspecto importante en la producción de café en Colombia, y puede tener un impacto en la oferta y la demanda del mercado de este grano. Así pues es necesario analizar cómo los productores de café en Colombia han adoptado prácticas agronómicas y de procesamiento que mejoran la calidad del café y garantizan la sostenibilidad de la producción, entre las que se incluyen la selección de variedades de café de alta calidad, el manejo adecuado de la planta de café, la recolección en el punto óptimo de madurez, el procesamiento

adecuado de los granos y el control de la calidad en todas las etapas del proceso (Muñoz-Belalcázar et al., 2021). El sector cafetero ha sido escenario de transformaciones y adaptaciones, buscando un objetivo: ser competitivos, pero siendo sostenibles. Las dinámicas cambiantes, el aumento de la competencia a nivel internacional, han sido motor natural para el surgimiento de estándares internacionales para las buenas prácticas de cultivo del grano, lo que a su vez ha sentado las bases para la creación de regulaciones y certificaciones internacionales cuya meta es satisfacer las demandas del mercado a nivel nacional e internacional (International Labour Organization, 2022).

Ahora bien, en cuanto a la sostenibilidad de la producción cafetera, en el sector se ha buscado que los productores implementen certificaciones de sostenibilidad como Fairtrade, Rainforest Alliance y UTZ¹, para demostrar su compromiso con prácticas agrícolas sostenibles y responsables. Estas certificaciones, además de ser un factor diferenciador, garantizan que el café es producido de manera socialmente justa, ambientalmente sostenible y económicamente viable (Oviedo-Celis y Castro-Escobar, 2021).

4.3 Inteligencia de negocios

La toma de decisiones basadas en datos se ha convertido en una necesidad determinante para los negocios, ya que permite a las organizaciones tomar decisiones más informadas y eficientes. En ese sentido la inteligencia de negocios (BI, por sus siglas en inglés) se constituye como una herramienta que ayuda a las organizaciones a recopilar, analizar y visualizar datos para mejorar la toma de decisiones (Shafi & Ahmad, 2014).

A medida que la toma de decisiones basadas en datos se vuelve esencial en el ámbito empresarial, no sólo en términos financieros sino en todos los niveles, se incrementa la necesidad de soluciones de BI. Diversos autores han sintetizado el significado de BI,

¹ UTZ: proviene de UTZ kapeh. Significa “buen café” en idioma maya guatemalteco (Rainforest Alliance, 2021).

ofreciendo una descripción concisa del concepto, por lo que se puede resumir como lo siguiente: La inteligencia de negocios es una herramienta, arquitectura, tecnología o sistema a través del cual información que se encuentra en estado crudo o sin tratar, puede ser organizada, estructurada y transformada, para transmitir información y generar conocimiento que facilite la toma de decisiones para quien lo utiliza, en últimas para organizaciones, empresarios o compañías que requieren esta información (Abu-Alsondos, 2023; Azevedo y Santos, 2009; Joyanes, 2019; Shafii y Ahmad, 2014)

A través de la BI se puede procesar información de cualquier tipo, como datos financieros, datos de producción, datos de ventas, etc., y estos a su vez, ser usados como fuente para la gestión de información clave de las organizaciones y como herramienta fundamental para el análisis de datos empresariales.

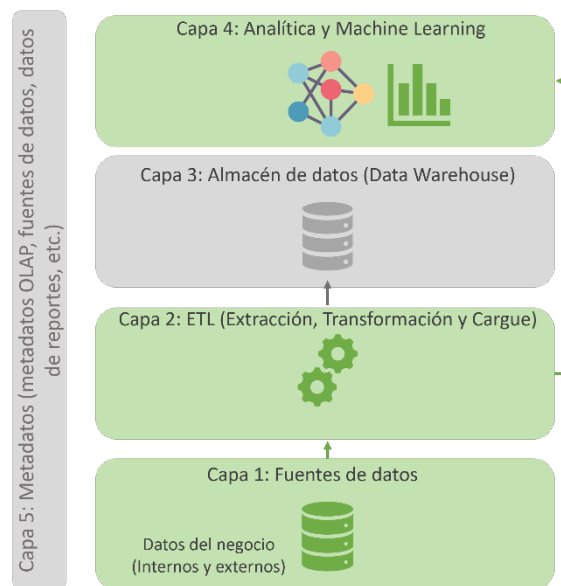
4.3.1 Modelo de inteligencia de negocios

Los modelos de inteligencia de negocios integran datos operativos y datos históricos con herramientas de analítica con el fin de presentar información valiosa para los tomadores de decisiones y planificadores dentro de las organizaciones (Shafi Reshi et al., 2014).

Se han realizado estudios previos, donde se hace uso de diversos modelos para las implementaciones de BI (Ereth y Baars, 2011; Khatibi et al., 2020; Ong et al., 2011; Rayed, 2019; Watson, 2009). Estos estudios, particularmente muestran que los modelos de inteligencia de negocios están conformados por distintas capas o componentes, los cuales tienen como finalidad guiar a la implementación del BI. No obstante, a pesar de que en las investigaciones mencionadas anteriormente se presentan distintas arquitecturas, se identificaron que las capas de extracción, transformación, almacenamiento, y reporte, hacen parte de los elementos esenciales para llevar a cabo una implementación exitosa de BI.

Estas capas permiten a las organizaciones recopilar, limpiar, organizar y analizar datos de manera eficiente. Teniendo en cuenta lo anterior, en la Figura 6 se presenta la propuesta para el presente trabajo, del modelo de capas de inteligencia de negocios. El modelo de capas de BI propuesto también incluye una capa de analítica y machine learning. En esta última se realizará la construcción y evaluación del modelo que busca predecir la demanda de exportación de café verde en Colombia. Cabe aclarar que en la arquitectura señalada en la Figura 6 también hace referencia a dos capas en gris, las cuales no serán profundizadas en este trabajo.

Figura 6 Arquitectura de capas propuesta



Fuente: Elaboración propia. Basado en Ong et al. (2011)

4.3.1.1 Capa de fuentes de datos

En esta capa se identifican y comprenden las fuentes de datos, esto implica estudiar a fondo los datos que se utilizarán para el modelado. Esta etapa es fundamental para realizar con mayor certidumbre la preparación de los datos (capa ETL) (Ong et al., 2011). Para entender los datos es necesario acceder a estos y realizar análisis exploratorios, que se suelen hacer mediante tablas y gráficos, así se podrá revisar la calidad de estos y describir los

resultados. Las fuentes de datos definidas para el presente trabajo son la FNC y el Banco de la República.

4.3.1.2 Capa de ETL

En los procesos de inteligencia de negocios, en muchas ocasiones es necesario integrar datos de diferentes fuentes, o simplemente datos heterogéneos. Por tanto, se hace necesario aplicar procesos de limpieza, corrección, tratamiento de datos faltantes o nulos, transformación y/o preparación de los datos con el fin de que éstos puedan ser usados en la capa de analítica y modelamiento con ML (Khatibi et al., 2020; Ong et al., 2011). Los datos, luego de ser tratados y limpiados, son transformados. La transformación implica que los datos sean convertidos en reglas del negocio y en formatos consistentes para el reporte y análisis que se pretende realizar. El cargue es la última fase del proceso de ETL, en donde como su nombre lo indica, los datos son cargados a un repositorio listos para ser analizados.

4.3.1.3 Capa de almacén de datos

Esta capa comprende el almacén donde se preservan todos los datos de las fuentes determinadas, usando un modelo multidimensional (Khatibi et al., 2020; Ong et al., 2011). Estas dimensiones hacen referencia a las perspectivas sobre las que se representan los hechos, a su vez, los hechos usualmente son medidas o valores numéricos y son referenciados mediante elementos de dimensión (Khatibi et al., 2020). En un almacén de datos, también es posible guardar datos históricos por lo que como buena práctica es común establecer ventanas de tiempo en donde se llevan a cabo actualizaciones de dichos datos (Ong et al., 2011).

4.3.1.4 Capa de analítica y machine learning

En esta capa se realiza la construcción y evaluación de los modelos predictivos del café verde en Colombia usando ML. A través de la exploración y aplicación de diversas técnicas

de ML, se buscará extraer patrones significativos de los datos históricos. Con la construcción de dichos modelos no solo se pretende obtener predicciones más precisas sobre la demanda futura, sino que también permitirán una comprensión profunda de los factores que influyen en las fluctuaciones de la demanda. Este análisis detallado, respaldado por las capacidades predictivas del aprendizaje automático, potenciará la toma de decisiones estratégicas en la gestión de la cadena de suministro de café verde en el contexto colombiano.

4.3.1.5 Capa de metadatos

Según Ong et al. (2011), esta es una de las capas más importantes y crítica dentro de una arquitectura de BI, ya que es donde los metadatos, las fuentes de datos, los datos procesados por ETL, etc., son almacenados y en donde pueden ser consultados. Así mismo se señala, que esta capa hace parte de la administración y gobernanza de los datos, por tanto, se resalta la importancia del manejo y la seguridad de estos con el fin de garantizar la escalabilidad de los procesos de BI (Ereth y Baars, 2011).

4.3.2 Implementaciones de inteligencia de negocios en el sector del cafetero

A la luz de la producción científica relacionada con la inteligencia de negocios, aplicada a predicción de demanda en general y en el sector cafetero, se realizó una revisión de la literatura en bases de datos con el fin de identificar tendencias, métodos, variables relevantes y hallazgos clave obtenidos por investigaciones previas. Para esta revisión se seleccionó la base de datos Web of Science², en esta base de datos se llevaron a cabo consultas mediante el uso de una ecuación de búsqueda, a partir de la cual se refinaron los resultados. En primera instancia se realizó la búsqueda utilizando la ecuación mostrada en la Figura 7, tal como se

² Web of Science: plataforma con varias bases de datos de búsqueda bibliográfica. Tomado de <https://clarivate.libguides.com/webofscienceplatform>

puede observar, esta primera revisión corresponde a un análisis general de la producción científica.

Figura 7 Ecuación de búsqueda Web of Science

```
TS=("coffee" OR "coffee invent*" OR "coffee export*" OR "coffee export* predic*" OR "coffee produc*" OR "coffee beans" OR "commodit*" OR "commodi* demand" OR "coffee demand" OR "demand forecast*" OR "demand prediction" OR "demand estimation") AND TS=("machine learning" OR "artificial intelligence" OR "analytics" OR "business intelligence") AND TS=("coffee pric*" OR "green coffee" OR "arabica" OR "robusta" OR "supply chain" OR "agribusiness" OR "agricul* industr*")
```

Fuente: Elaboración propia

La búsqueda arrojó 761 resultados, así pues, tras realizar la revisión general, se procedió a utilizar las herramientas de filtros rápidos de Web of Science, entre los cuales están la fecha de publicaciones (2018-2023), tipo de documento, categorías específicas como investigación de operaciones, administración y ciencias de la computación, entre otros. Este proceso redujo los resultados a 155, que se evaluaron mediante una lectura de resúmenes para identificar la relevancia al tema. A continuación, se presenta un resumen de la revisión bibliográfica:

La mayoría de los artículos se publicaron en 2022 (7 documentos), mostrando una tendencia ascendente hasta ese año, mientras que en 2023 se registró solo un documento.

Entre las categorías donde se muestran los artículos, está administración, ciencias de los alimentos, agricultura, agronomía y ciencias ambientales, el resumen de los artículos se puede observar en la Tabla 2.

Tabla 2 Resumen de revisión bibliográfica

Autor	Métodos utilizados	Variables	Resultados encontrados
Kittichotsawat et al., (2022)	Los autores hacen tal comparativo y utilizaron éstas técnicas para predecir la producción anual en la siembra de café. El modelo de ML que aplican es Redes	Para la predicción de la producción de la siembra de café, incluyeron variables como el área cultivada, lluvias,	Los autores evaluaron el desempeño del modelo de predicción de la siembra a través del coeficiente de correlación (R2) y la raíz del error cuadrático medio (RMSE, por sus siglas en

Autor	Métodos utilizados	Variables	Resultados encontrados
	<p>Neuronales Artificiales (ANN, por sus siglas en inglés). Para implementar el modelo ANN, los autores realizan la partición de los datos en 3 grupos: 70% usado para entrenamiento, 15% para prueba y otro 15% para validación, esto lo hizo con el fin de evitar el sobreajuste (Overfitting, por su nombre en inglés).</p>	<p>humedad relativa, las zonas productivas, temperaturas mínimas y máximas. La variable de salida o respuesta fue el rendimiento de productividad anual. Utilizaron datos de 5 años de 2004 a 2018 (180 meses).</p>	<p>inglés). Encontraron que el R2 y el RMSE del modelo con ANN presentó mejor desempeño que ARIMA para la estimación de la producción de café arábica anual.</p>
<p>Deina et al., (2022)</p>	<p>Utilizaron algoritmos de ML para estimar los precios del café Arábica y Robusta. Para ello, utilizaron los siguientes modelos: Suavización exponencial (ES, por sus siglas en inglés), Autoregresión (AR, por sus siglas en inglés), ARIMA, Perceptrón multicapa (MLP, por sus siglas en inglés), redes neuronales como Máquinas de Aprendizaje Extremo (ELM, por sus siglas en inglés).</p>	<p>Los autores usaron como variable el precio promedio mensual del café Arábica y Robusta. Los datos para esta investigación correspondían a precios entre 1996-2018 y 2001-2018 respectivamente.</p>	<p>Se usó el error cuadrático medio (MSE, por sus siglas en inglés), el error medio absoluto (MAE, por sus siglas en inglés) y el porcentaje de error medio absoluto (MAPE, por sus siglas en inglés). Basados en estos indicadores, Deina et al., (2022) determinaron que, en la predicción de los precios del café Arábica y Robusta, el modelo ELM presentó mejor desempeño frente a las demás técnicas consideradas. Además, estos autores resaltan que el modelo ELM tiene un menor costo computacional en comparación con otros modelos como el MLP.</p>
<p>Bacci et al., (2019)</p>	<p>Se analizaron métodos para predecir la demanda interna de café en Brasil, mediante el uso de Análisis Factorial e</p>	<p>Proponen el uso de la variable de consumo interno de café (en miles de sacos de</p>	<p>Los autores también usaron las pruebas Diebold-Mariano (pruebas DM) para evaluar las diferencias significativas entre</p>

Autor	Métodos utilizados	Variables	Resultados encontrados
	Intersección de Límites Normales (FA-NBI, por sus siglas en inglés) y Diseño de Experimentos Mixtos (DOE-M, por sus siglas en inglés).	60kg) para predecir la demanda interna de café en Brasil para 2006-2007 y 2017-2018.	los errores. En el caso de Bacci et al., (2019) las pruebas DM confirmaron la mejora significativa de la precisión (rendimiento) del FA-NBI en comparación con los métodos de ponderación individual y otros métodos (ver Sección 4.3.2.1). El análisis demostró que el FA-NBI superó a los métodos individuales (DES, WM, ARIMA(1,1,1), ARIMA(2,2,3)) en términos de diversas métricas de rendimiento (MAE, MSE, RMSE, MAPE, entre otras). La aplicabilidad del modelo FA-NBI fue probada en series de tiempo reales y simuladas de producción y exportaciones de café de Brasil.
Mejía y Rincón, (2021)	Haciendo uso de las técnicas de grandes volúmenes de datos (Big Data, por su nombre inglés) buscaron determinar la relación entre las exportaciones de café, el precio del grano en mercados internacionales y la producción de café.	Se usaron variables como el volumen mensual de exportaciones y el precio, tanto interno como externo, de estas. Estas variables fueron analizadas con el fin de determinar patrones de comportamiento.	Los autores encontraron una relación directa entre las exportaciones y el precio del café interno y externo. Es decir, si las exportaciones aumentan, también lo hace el precio del café.

Autor	Métodos utilizados	Variables	Resultados encontrados
Jiménez-Varón y Tabares-Soto, (2018)	Los autores analizaron el comportamiento de la demanda y oferta del café verde utilizando modelos estadísticos	En su investigación estudiaron el comportamiento de variables como la demanda, oferta y stock de café verde en Colombia, a partir de las cuales se planteó un modelo mediante el uso de ecuaciones diferenciales ordinarias	Se encontró que la variable de producción tiene gran influencia en el comportamiento del mercado, esto debido principalmente a la capacidad instalada que se requiere para su gestión. Así mismo, encontró que la demanda de café en Colombia es altamente regida por la demanda internacional

Fuente: Elaboración propia

4.3.2.1 Conclusiones de la revisión bibliográfica

La revisión bibliográfica realizada resalta la importancia de investigar la demanda de café. Aunque se ha observado un interés creciente en la investigación, centrado en ejes temáticos como la inteligencia de negocios y modelos predictivos asociados al café, es notorio que estos esfuerzos se han enfocado predominantemente en aspectos productivos, calidad y predicción de cultivos, abordando, en su mayoría, el primer eslabón de la cadena agroindustrial del café. Sin embargo, se evidencia una carencia de estudios que analicen la demanda específica del café verde, siendo aún más escasos los realizados en el contexto colombiano.

La revisión también destaca la limitada aplicación de técnicas de machine learning en el eslabón de la cadena productiva del café relacionado con las exportaciones. Este vacío de conocimiento crea una oportunidad para profundizar en esta área y capitalizar el potencial de las herramientas de ML en la predicción de la demanda de café verde en Colombia. Frente a estas perspectivas, la presente investigación se plantea como un esfuerzo para llenar este vacío de conocimiento, proporcionando una aproximación a la inclusión de modelo de ML en

el marco de una arquitectura de BI y arrojando luz sobre la demanda de café verde en Colombia mediante la aplicación de técnicas avanzadas de aprendizaje automático.

4.4 Modelos de predicción

Se entiende como pronóstico el proceso de estimar eventos futuros a partir de datos históricos y modelos matemáticos (Chopra y Meindl, 2013; Heizer & Render, 2012). La idoneidad de estos modelos depende de diversos factores, incluyendo el horizonte temporal, la información histórica disponible y los patrones de comportamiento de la demanda real (Chopra y Meindl, 2013; Heizer y Render, 2012). Este proceso involucra la evaluación del error de pronóstico, indicador del desempeño del método utilizado, comparándolo con la información histórica disponible (Heizer et al., 2009). Existen dos categorías principales de métodos de pronóstico: cualitativos y cuantitativos.

Los métodos de pronóstico cualitativos se usan generalmente para los casos en los que no es posible cuantificar los datos o no son suficientes. Para estas situaciones se hace uso de diferentes metodologías como son: el juicio de expertos, el método Delphi, la modelación de escenarios y enfoques intuitivos (Villarreal, 2016).

Los métodos cuantitativos se utilizan en diversas áreas para predecir tendencias futuras, estos requieren, para su desarrollo, de modelos matemáticos y su implementación se lleva a cabo a través de herramientas informáticas. El insumo para la implementación de estos modelos son los datos históricos de una o más variables de interés (Sanders y Manrodt, 2003).

La presente investigación se enfoca en la aplicación de métodos cuantitativos, al evaluar modelos de predicción de la demanda de café verde en Colombia mediante algoritmos de aprendizaje automático. Estos métodos se basan en aprender de patrones en los datos históricos disponibles para una predicción, partiendo del supuesto de que la variable que

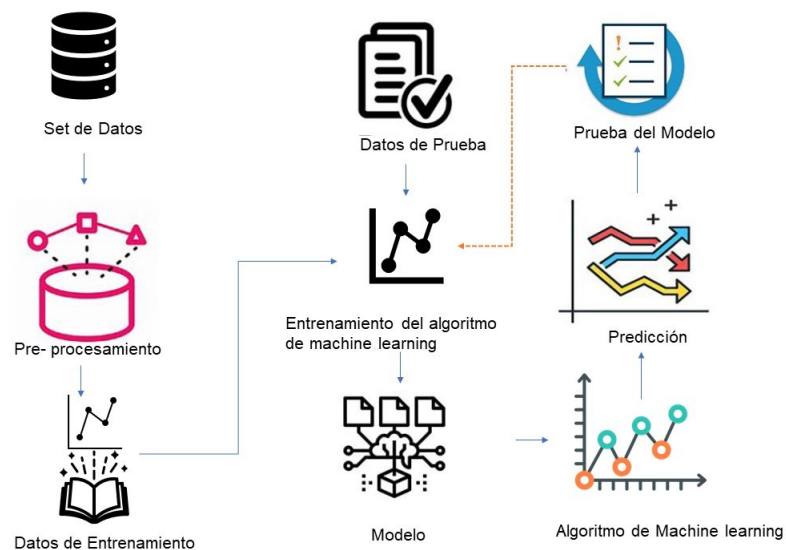
tratamos de pronosticar en este caso las exportaciones exhibe una relación de causa y efecto con una o más variables (Villarreal, 2016).

4.4.1 Modelos de aprendizaje automático

El aprendizaje automático también conocido como machine learning, es un área de la Inteligencia Artificial, que proporciona algoritmos para clasificar y predecir variables de interés (Joyanes, 2019). Los algoritmos de ML se pueden clasificar en supervisados y no supervisados.

En el aprendizaje supervisado, utilizado en esta investigación, los datos o conjunto de datos que se usan como entrada al sistema poseen una etiqueta y partiendo de este grupo de datos se pueden utilizar diferentes algoritmos para educar el modelo y obtener las salidas deseadas (Joyanes, 2019). La Figura 8 corresponde a las etapas que se llevan a cabo para implementar un modelo de ML para predicción.

Figura 8 Etapas para la implementación de un modelo de Machine learning



Fuente: Elaboración propia, adaptado de Kittichotsatsawat et al., (2022)

Como se puede observar en la Figura 8, la implementación de modelos de Machine learning, sigue la arquitectura de capas de BI propuesta en la sección anterior (Figura 6). La primera fase de la implementación es la construcción y preprocesamiento del conjunto de datos para que estos sean adecuados para el aprendizaje automático. La siguiente fase es la partición del conjunto de datos en datos de entrenamiento y prueba. Los datos de entrenamiento se utilizan para el aprendizaje del modelo, mientras que los datos de prueba se utilizan para evaluar su precisión. Para esto, se eligen los algoritmos supervisados, los cuales se entrenan con datos etiquetados, que contienen los valores correctos para la variable que se desea predecir. El entrenamiento del modelo implica ajustar los parámetros del algoritmo para que haga predicciones lo más precisas posible (Benos et al., 2021).

En la fase de evaluación, se utiliza un conjunto de prueba o a través de validación cruzada para verificar la precisión del modelo (Ghatak, 2017). Si los resultados son insatisfactorios, se pueden revisar posibles problemas de sobreajuste o subajuste. Es importante destacar que estos pasos adicionales, como la separación de datos, el entrenamiento del algoritmo, la prueba del modelo y las predicciones, complementan la arquitectura general de inteligencia de negocios, además de ser esenciales para garantizar que los modelos de machine learning sean precisos y útiles.

4.4.2 Selección de variables

La selección de las variables en ML es el proceso de reducir las variables o características del conjunto de datos con el fin de generar un mayor rendimiento de predicción de los modelos (Kuhn y Johnson, 2013). Según el tipo de variables que se estén analizando se pueden utilizar diferentes métodos para la selección, los cuales se clasifican en supervisados y no supervisados dependiendo si tienen en cuenta o no la variable de salida (Kuhn & Johnson, 2013). En este trabajo se usan métodos supervisados, porque la selección de variables se

hará por método de filtro considerando su relación con la variable objetivo, teniendo en cuenta que las variables de entrada y de salida que se utilizarán son variables continuas de tipo numérico, el método estadístico más adecuado es el análisis de correlación, mediante el cual se seleccionan las variables que están altamente correlacionadas con la variable objetivo, pero débilmente correlacionadas entre sí, con el objetivo de reducir la multicolinealidad (Kuhn y Johnson, 2013).

4.4.3 Datos de entrenamiento y prueba

La evaluación de los algoritmos de ML requiere analizar no solo su rendimiento con los patrones de aprendizaje, sino también su comportamiento ante conjuntos de datos no utilizados en el entrenamiento (Santana, 2006). Para ello, es necesario disponer de dos conjuntos de datos: el conjunto de entrenamiento y el conjunto de prueba (Santana, 2006). La división del conjunto de datos principal en estas dos categorías permite primero entrenar el algoritmo con el conjunto de entrenamiento y luego aplicar el modelo entrenado al conjunto de prueba, que consta de datos no vistos. Al aplicar el modelo entrenado a los datos de prueba, se evalúa su capacidad para generalizar y realizar predicciones precisas en datos no vistos (Ghatak, 2017). En la literatura son encontrados distintos métodos para realizar la partición de los datos. Los métodos más comunes son la partición fija o aleatoria, en la que se asigna un porcentaje del conjunto de datos al conjunto de entrenamiento y el resto al conjunto de prueba (por ejemplo, 80%-20%, respectivamente); y la validación cruzada, en la que el conjunto de datos se divide en k partes, de forma que cada parte se utiliza una vez como conjunto de prueba, mientras que las otras $k - 1$ partes se utilizan como conjunto de entrenamiento (Korjus et al., 2016).

Dicho método consiste en dividir el conjunto de datos en un número determinado de subconjuntos (k partes), donde $k = 10$ suele ser el número encontrado en la literatura a través

de prueba-error que mejor desempeño presenta (Korjus et al., 2016). De esta forma, cada subconjunto actúa como conjunto de prueba mientras el modelo se entrena con los $k - 1$ restantes. Este proceso se repite varias veces, calculando métricas de rendimiento promediadas. La validación cruzada ayuda a obtener una evaluación más robusta del modelo al utilizar diferentes combinaciones de datos de entrenamiento y prueba; por tanto, en este trabajo se selecciona este método para realizar el entrenamiento y validación del modelo (Contreras-Bravo et al., 2022; Korjus et al., 2016).

4.4.4 Sobreajuste y subajuste

El sobreajuste (*overfitting*) y subajuste (*underfitting*) son conceptos clave al construir modelos de ML. El subajuste ocurre cuando el modelo no puede explicar los datos, por tanto, se observa su incapacidad para captar las tendencias en los datos; en general, este fenómeno suele estar asociado a modelos demasiado simples (Korjus et al., 2016). Por su parte, el sobreajuste se presenta cuando el modelo se ajusta excesivamente a los datos, capturando incluso el ruido de los datos y variaciones no relevantes; y a menudo es asociado con una gran cantidad de parámetros o características en los modelos (Ghatak, 2017). Para abordar este riesgo, en este trabajo, se utiliza validación cruzada ya que, en cada iteración, el modelo se ajusta a diferentes subconjuntos de datos, lo que ayuda a evitar que se sobreajuste a patrones específicos (*overfitting*) o que sea demasiado simple y no capture la complejidad de los datos (*underfitting*) y contribuyendo así a mejorar la capacidad de generalización del modelo construido (Ghatak, 2017; Korjus et al., 2016).

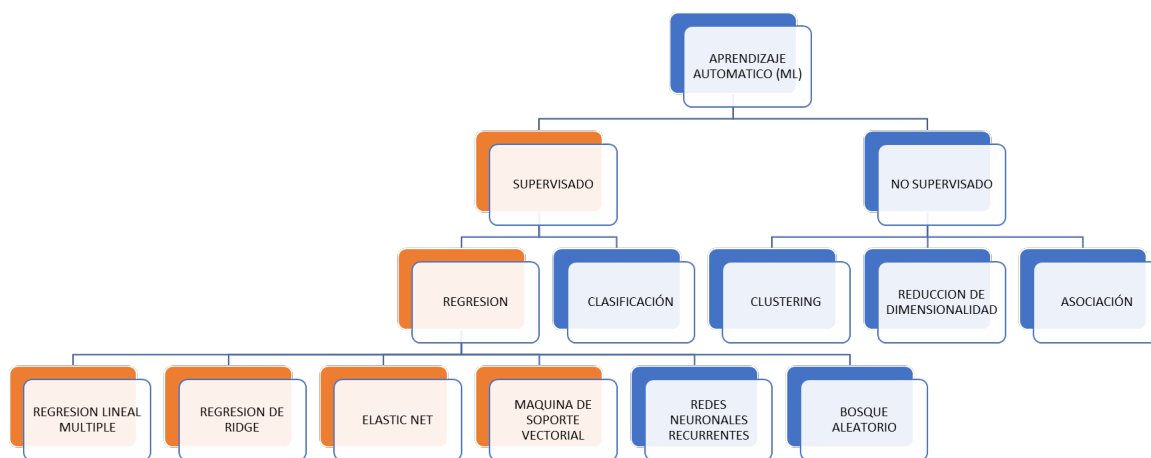
4.4.5 Algoritmos utilizados para la predicción de la demanda

En el proceso de predecir la demanda externa del café verde en Colombia, se lleva a cabo la construcción y evaluación de algoritmos de ML (regresión lineal, regresión de Ridge, Red Elástica y Máquinas de Soporte Vectorial); estos algoritmos desempeñan un papel

fundamental en la predicción de la variable independiente (salida), basándose en una o más covariables que actúan como insumos para el modelo. La elección de estos algoritmos está respaldada por la revisión de la literatura previamente realizada.

En la Figura 9 se muestran los algoritmos de regresión seleccionados para la construcción de los modelos de ML, bajo aprendizaje supervisado.

Figura 9 Algoritmos de ML seleccionados



Fuente: Elaboración propia a partir de Ghatak, 2017; Kuhn y Johnson, (2013)

4.4.5.1 Regresión lineal

La regresión con una sola variable x es conocida como regresión simple, y es útil para predecir una variable respuesta con una sola entrada. Es un tipo de algoritmo de aprendizaje supervisado, lo cual significa que los valores de salida son conocidos en el entrenamiento de los datos (Ghatak, 2017). Por otro lado, cuando existen 2 o más covariables, se utiliza el modelo de regresión lineal múltiple, cuya ecuación se representa de la siguiente manera:

$$Y_t = b_0 + b_1X_{1,t} + b_2X_{2,t} + \dots + b_kX_{k,t} + \varepsilon_t \quad (1)$$

Dónde y es la variable para pronosticar y son $X_1 \dots X_k$, las variables predictoras. Los coeficientes $b_1 \dots b_k$, miden el efecto de cada predictor después de tener en cuenta los efectos

de todos los demás predictores del modelo. Así, los coeficientes miden los efectos marginales de las variables predictoras y ε_t corresponde al término del erro (Ghatak, 2017).

4.4.5.2 Regresión de Ridge

La regresión de Ridge es una técnica de regularización utilizada para crear un modelo mejor y más preciso (Moreno y Cristiano, 2021). Esta técnica impone una penalización al tamaño de los coeficientes de la relación lineal entre las variables predictoras y la variable objetivo. Los coeficientes calculados minimizan la suma de los cuadrados de los residuos penalizada.

$$RSS_{ridge} = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (2)$$

Donde λ es el parámetro que controla el grado de penalización: cuanto mayor es este menor serán los coeficientes, resultando más robustos a la colinealidad. La estimación de un buen valor para λ , es fundamental, por lo general en ML, se realiza mediante validación cruzada (Ghatak, 2017; Moreno y Cristiano, 2021).

4.4.5.3 Red elástica

La Red Elástica (*Elastic Net*, por su nombre en inglés) es un método de regularización utilizado en estadísticas y aprendizaje automático, especialmente en problemas de regresión. Combina dos técnicas de regularización, Lasso y Ridge, para abordar problemas de multicolinealidad y para seleccionar automáticamente características importantes en un conjunto de datos (Von Lücken, 2021). La elasticidad en *Elastic Net* se refiere a la capacidad del método para ajustar los coeficientes de las características de manera más flexible en comparación con Lasso o Ridge por sí solos. Al ajustar los hiperparámetros³ de *Elastic Net*, se puede controlar la proporción de penalización, lo que permite equilibrar la selección de características y la reducción de la magnitud de los coeficientes (Hans, 2011).

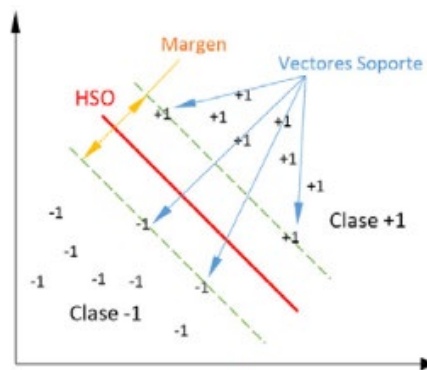
³ Hiperparámetro: Son variables externas al modelo de machine learning, utilizadas para controlar el proceso de entrenamiento (Ghatak, 2017).

4.4.5.4 Máquinas de soporte vectorial SVM

Las Maquinas de soporte vectorial (SVM, por sus siglas en inglés), son otro tipo de algoritmo de aprendizaje supervisado que puede ser usado en problemas de regresión (SVR, por sus siglas en inglés) o clasificación (SVC, por sus siglas en inglés). El objetivo de este algoritmo es construir un hiperplano⁴ que divida un conjunto de datos en dos clases o valores de salida tal como se puede ver en la Figura 10 (Sánchez, 2016). Este proceso se realiza a través de la manipulación de diversos parámetros, entre los cuales se incluyen:

- Núcleo (kernel): funciones utilizadas para el reconocimiento de patrones intrínsecos en los datos, pudiendo ser lineales o no lineales.
- 'C': parámetro de regularización que controla la penalización por errores en la función de regresión.
- Sigma (σ): controla la flexibilidad del modelo. Un valor de sigma bajo proporciona un modelo más suave y generalizado, mientras que un valor alto puede llevar a un modelo más ajustado y específico para los datos de entrenamiento.

Figura 10 Hiperplano de separación óptimo construido con SVM



Fuente: Tomado de González et al., (2017)

R cuenta con el paquete e1071 el cual soporta la implementación de este algoritmo. De tal forma, que en este trabajo se emplea el algoritmo SVM con una función (núcleo) de base radial

⁴ Hiperplano: en términos de máquinas de soporte vectorial, es un plano que separa y clasifica un conjunto de datos (*Máquina de Soporte vectorial SVM – numerentur.org, s. f.*).

(RBF, por sus siglas en inglés) y la selección óptima de hiperparámetros se basa en métricas de rendimiento como la raíz del error cuadrático medio (RMSE) (Cuevas et al., 2019).

4.4.6 Métricas para comparar modelos y evaluar su rendimiento

Evaluar la exactitud de los pronósticos resulta fundamental independientemente de la técnica utilizada. El error aleatorio se debe en parte a factores no predecibles que afectan el resultado y lo alejan de la demanda real. Aunque los especialistas en pronósticos intentan reducir los errores mediante la elección de modelos adecuados, es imposible eliminar completamente todas las fuentes de error (Krajewski et al., 2013). A continuación, se muestran las métricas seleccionadas en este trabajo para evaluar los modelos construidos.

4.4.6.1 Error absoluto medio

El error absoluto medio (MAE, por sus siglas en inglés) es un indicador utilizado en estadística y aprendizaje automático para medir la magnitud promedio de los errores entre los valores previstos y reales. MAE se calcula como el promedio de las diferencias absolutas entre los valores previstos y los valores reales (Deina et al., 2022). Su definición viene dada de la siguiente forma:

$$MAE = \frac{1}{n} * \sum |A_t - F_t| \quad (3)$$

t = número de periodos

A_t = demanda real del periodo t

F_t = pronóstico del periodo t

n = número total de periodos

$| |$ = símbolo de valor absoluto.

4.4.6.2 Coeficiente de determinación R^2

El coeficiente de determinación (RSQ por sus siglas en inglés) es una medida estadística que representa la proporción de la varianza de una variable dependiente que se explica por una o varias variables independientes en un modelo de regresión.

El valor de R^2 de un modelo, en RStudio es:

```
summary(model) $r.squared
```

Un valor cercano a 1 indica que el modelo explica la mayor parte de la varianza de la variable objetivo. Un bajo valor, cercano a 0, implica que es necesario incluir otras características importantes en el modelo (Ghatak, 2017).

4.4.6.3 Raíz de error cuadrático medio

Este indicador mide la diferencia media entre los valores previstos por el modelo predictivo y los valores reales.

El RMSE (por sus siglas en inglés) se calcula usando la siguiente fórmula (Kittichotsatsawat et al., 2022):

$$RMSE = \sqrt{\frac{1}{n} \sum (A_t - F_t)^2} \quad (4)$$

donde:

A_t = demanda real del periodo t

F_t = pronóstico del periodo t

n = número total de periodos

5 Hipótesis

En concordancia con los objetivos y la revisión de la literatura, se plantea la siguiente hipótesis:

Incluir modelos de predicción de demanda externa en una arquitectura de inteligencia de negocios, contribuye a la toma de decisiones informadas por parte de los actores que participan en la cadena de suministro del café verde en Colombia.

6 Fuentes de datos y variables

En el marco del presente trabajo se analizan múltiples variables obtenidas de fuentes de datos abiertos de la Federación Nacional de Cafeteros de Colombia (FNC), entidad encargada de establecer los lineamientos de calidad y exportaciones de café en Colombia (*Cafés Suaves - Federación Nacional de Cafeteros*, 2019). El conjunto de datos contiene las siguientes covariables: producción de café verde dado en miles de sacos de 60 kg, el precio exdock (precio del café verde internacional puesto en puerto), el precio (OIC) café suave colombiano, el precio indicativo compuesto y la tasa de cambio representativa del mercado (TRM). Así mismo como variable respuesta se tiene las exportaciones de café verde en miles de sacos de 60 kg. Estas variables, al ser analizadas en conjunto, arrojan una visión integral de los factores que influyen en la demanda externa de café verde en Colombia y sirven como base fundamental para la construcción del modelo predictivo.

6.1 Fuentes de datos

Una de las fuentes de datos definidas para el presente trabajo es la Federación Nacional de Cafeteros de Colombia (FNC). La FNC recopila y centraliza los datos de producción y exportación de café verde en Colombia, los cuales son abiertos al público. También se encarga de consolidar información sobre el sector cafetero, como los precios internacionales del café, los países a donde se exporta el café colombiano, los precios indicativos de la Organización Internacional del Café, entre otros indicadores. Esta información se descarga de la página web de la entidad en archivos de Excel, de los cuales se extraen las variables de estudio.

En la Figura 11 se puede visualizar los datos correspondientes a datos históricos de exportaciones colombianas dados en miles de sacos de 60 kg de café verde equivalente obtenidos de la FNC.

Figura 11 Volumen de exportaciones colombianas de café

MES	Total Exportaciones
ene-00	517
feb-00	642
mar-00	404
abr-00	731
may-00	615
jun-00	869
jul-00	721
ago-00	938
sep-00	674
oct-00	831
nov-00	1200
dic-00	1064
ene-01	651
feb-01	838
mar-01	610

Fuente. Información de base de datos obtenida de Estadísticas cafeteras - Federación Nacional de Cafeteros de Colombia, (s/f)

El Banco de la República, es la segunda fuente de datos que hace parte del presente trabajo, a partir del cual se obtienen los valores históricos de la tasa representativa del mercado (TRM). Esta variable es parte del conjunto de datos objeto de análisis, debido a que los precios internacionales del café se expresan en dólares estadounidenses, por lo cual un dólar más caro se materializa en más pesos que pueden recibir los productores de café.

6.2 Variables

De acuerdo con la arquitectura de BI propuesta, se requiere de la especificación de variables de entrada y salida, para lo cual se seleccionan las mostradas en la Tabla 2. Estas variables se seleccionan a partir de la revisión de literatura realizada (Deina et al., 2022; Gil, 2012; Kittichotsatsawat et al., 2022; Mejía y Rincón, 2021; Monsalve, 2016). Las variables de entrada o variables independientes son utilizadas por los modelos de machine learning con el

fin de ofrecer información del pronóstico de la cantidad de café verde a exportar. Se utiliza una ventana de tiempo de los datos desde enero del 2000 hasta diciembre del 2022, con 276 registros para cada variable.

Tabla 3 Resumen de variables

Variable	Definición	Unidad de medida	Clasificación	Tipo	Escala	Fuente
Cantidad de café verde exportado	Cantidad de sacos de 60 kg equivalentes exportados.	Miles de sacos de 60 Kg de café verde equivalente	Dependiente	Cuantitativa	Razón	Federación Nacional de Cafeteros de Colombia.
Precio Exdock	Precio externo del café colombiano (exdock)* - Promedio Mensual (centavos de dólar por libra de 453.6 gr de Café Excelso)	Centavos de dólar por libra de 453.6 gr de Café Excelso ⁵	Independiente	Cuantitativa	Razón	Federación Nacional de Cafeteros de Colombia
Tasa Representativa del Mercado (TRM)	Tasa Representativa del Mercado	Pesos colombianos por dólar estadounidense	Independiente	Cuantitativa	Razón	Banco de la República
Cantidad de café verde producido	Cantidad de sacos de 60 kg equivalentes producidos	Miles de sacos de 60 Kilogramos de café verde equivalente	Independiente	Cuantitativa	Razón	Federación Nacional de Cafeteros de Colombia.

⁵ Café Excelso: es un tipo de calidad del café almendra (café verde o pergamino después de trillar) que ha sido seleccionado por su alta calidad en las trilladoras de café (Federación Nacional de Cafeteros de Colombia, n.d.-a)

Precio OIC de café suave colombiano	Precio Promedio Mensual para el café suave por libra.	OIC para el café suave	Centavos dólar de 453.6 gr de Café	de libra	Independiente	Cuantitativa	Razón	Federación Nacional de Cafeteros de Colombia
Precio del indicador compuesto OIC	Precio del indicador compuesto OIC	del por libra	Centavos dólar de 453.6 gr de Café	de libra	Independiente	Cuantitativa	Razón	Federación Nacional de Cafeteros de Colombia

Fuente: Elaboración propia.

7 Diseño metodológico

La presente investigación es un trabajo aplicado de los conceptos adquiridos para evaluar un modelo de inteligencia de negocios para el pronóstico de la demanda externa de café verde en Colombia utilizando técnicas de aprendizaje automático, a partir de datos abiertos disponibles en la FNC y el Banco de la República. Por lo anterior, podemos concluir que la investigación es cuantitativa, ya que se revisan variables cuantitativas mediante un análisis de datos numéricos y herramientas de analítica descriptiva. El horizonte temporal de los datos analizados en la investigación es del año 2000 al 2022. No se toma muestra, por lo que se utiliza toda la información disponible para la elaboración de un modelo de ML en el marco de una arquitectura BI.

Fases de la Investigación

7.1 Fase 1

Dando cumplimiento al objetivo específico número 1, se realiza una revisión de la literatura para guiar la construcción del modelo de predicción de la demanda externa de café verde en Colombia desde el enfoque de la inteligencia de negocios y cadenas de suministro, utilizando

como herramientas de búsqueda bases de datos como Scopus®, Web of Science®, entre otros, con el fin de identificar los factores y metodologías aplicadas en estudios relacionados a la predicción de la demanda del café verde.

Como parte de los resultados de esta primera fase, se determinan las etapas de implementación de un modelo de ML, se identifican las variables útiles para proponer un modelo en el contexto de interés, se proponen las técnicas o métodos a utilizar en el desarrollo del trabajo y se contextualiza el uso de estas técnicas en el escenario de BI y la cadena de suministro del café.

7.2 Fase 2

En esta fase, se caracterizan los factores económicos que influyen en la demanda de exportación de café verde en Colombia mediante la revisión y recopilación de variables clave de bases de datos, incluyendo la FNC y el Banco de la República, para obtener datos de producción, exportación y precios internacionales del café verde entre 2000 y 2022. Subsecuentemente, se lleva a cabo el proceso ETL y la construcción del modelo de datos utilizando Power BI. El resultado es un conjunto de datos listo para analizar y predecir las exportaciones de café verde. Para comprender mejor el contexto de estas variables, se realiza un análisis descriptivo y exploratorio, destacando tendencias y patrones que orientan la selección de los modelos de ML más apropiados, incluyendo la evaluación de la correlación entre las variables independientes y su influencia en la variable a predecir (exportaciones de café verde en Colombia).

7.3 Fase 3

Con el objetivo de construir y evaluar un modelo de predicción de la demanda de café verde en Colombia, se implementan diversos algoritmos de ML, seleccionados con base en la revisión bibliográfica realizada en la Fase 1. Después de validar los modelos construidos, se

lleva a cabo la evaluación del rendimiento, utilizando métodos como regresión lineal múltiple, Red Elástica y Máquinas de Soporte Vectorial. En la discusión de resultados, se realiza la elección del modelo considerando indicadores de error de pronóstico como RSME, R-cuadrado y MAE, a partir de los resultados obtenidos en la evaluación de los diferentes modelos y posteriormente se aborda como estos resultados pueden llegar al usuario final del modelo de BI, como información útil, para el proceso de toma de decisiones.

7.4 Fase 4

Durante esta fase se proponen las orientaciones metodológicas para la predicción de la demanda de café verde en Colombia bajo un enfoque de inteligencia de negocios utilizando el modelo seleccionado, dando alcance al cuarto objetivo específico del presente trabajo.

Finalmente, luego de realizadas las cuatro fases se presentan las conclusiones y trabajos futuros que se puede derivar de la presente investigación.

8 Trabajo de Campo

8.1 Proceso ETL, almacenamiento y modelo de datos

Para caracterizar los factores económicos que influyen en el comportamiento de la demanda del café verde en Colombia, se trabajan las dos primeras capas de la arquitectura propuesta en la Figura 6. Estas dos primeras capas corresponden a la extracción de fuente de datos y al proceso de ETL.

A partir de estos procesos, se realiza la construcción de la base de datos con la información recolectada de la FNC y del Banco de la República. Para ello, se crea una llave con la que se

permite el relacionamiento entre las distintas variables, esta llave corresponde a un ID asignado a cada registro que se tiene en la base de datos (ver Figura 12).

Figura 12 Visualización de primeras filas de la base de datos construida

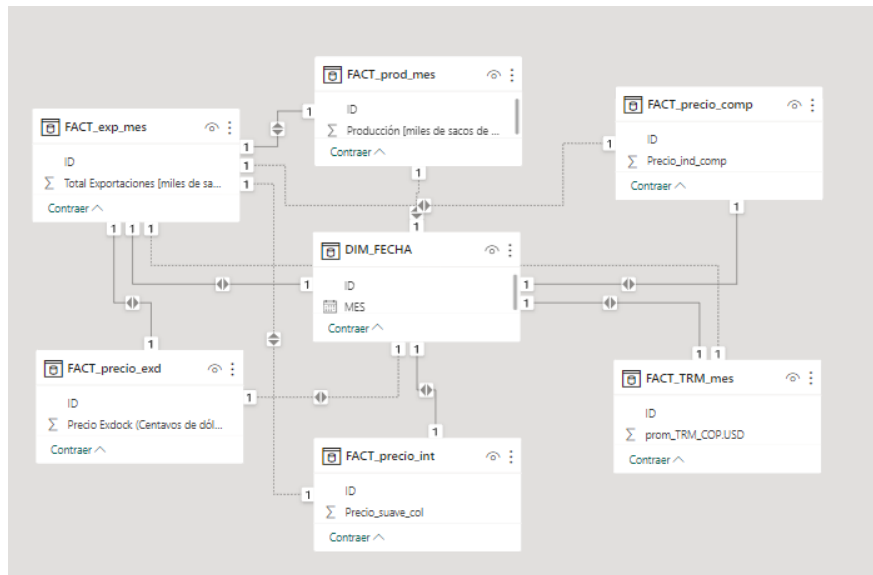
	A	B	C	D	E	F	G
1	MES	TRM_COP.USD	Precio_Exdock	Precio_ind_comp	Precio_suave_col	Produccion	Exportaciones
2	1/01/2000 0:00	1923.57	126.56	82.15	130.12	658	517
3	1/02/2000 0:00	1950.64	119.87	76.15	124.72	740	642
4	1/03/2000 0:00	1956.25	117.69	73.49	119.51	592	404
5	1/04/2000 0:00	1986.77	109.77	69.53	112.67	1055	731
6	1/05/2000 0:00	2055.69	111.84	69.22	110.31	1114	615
7	1/06/2000 0:00	2120.17	102.4	64.56	100.3	1092	869
8	1/07/2000 0:00	2161.34	103.78	64.09	101.67	811	721

Fuente: Elaboración propia

Posterior a la creación de la llave, se procede a realizar el relacionamiento con cada una de las variables que hacen parte de la base de datos, mediante Power Query en Power BI. En el proceso se crea una tabla de datos de dimensión correspondiente a la fecha, representada mes a mes, donde se han realizado el registro de cada una de las variables. Además de lo anterior, se crean seis tablas de hechos correspondientes a las variables de producción mensual de café verde, precio Exdock, precio OIC de café suave colombiano, precio del indicador compuesto OIC, la TRM y exportaciones de café verde (ver Figura 13).

En este trabajo, se utiliza el almacén de datos propio de Power BI, orientado a sujeto, el cual se enfoca en las entidades claves, por lo que no se aborda el análisis de la estructura temporal de los datos.

Figura 13 Construcción del modelo de datos



Fuente: Elaboración propia

Como parte de la transformación de los datos, se realiza la selección de la ventana de tiempo de los datos desde enero del año 2000 hasta diciembre del año 2022. No es necesario realizar imputación o reemplazo de datos faltantes.

8.2 Caracterización de factores económicos que influyen en el comportamiento de la demanda de exportación de café verde en Colombia

En la Figura 14 se puede observar el comportamiento de la producción y exportaciones de café verde presentan similar tendencia a lo largo de los años. En dado caso, se puede analizar que tienen un comportamiento proporcional ya sea hacia el aumento o disminución; también es evidenciable que la producción se encuentra por encima de los valores de exportación de café verde. Sin embargo, la diferencia entre los valores históricos entre ambas variables (producción y exportación) parece no ser muy amplia, esto sugiere una falta de preparación de inventarios de seguridad necesario para exportación, en este caso, es notable que se tienen valores máximos de inventarios de 1,400,000 sacos de café, mientras que como valores mínimos se tiene -324,000 sacos, lo que indica que se ha llegado al desabastecimiento para

exportación. Dicho desabastecimiento se hace evidente en los años 2008, 2009, 2021 y 2022, donde la producción fue alcanzada por la demanda externa (exportación).

En el caso particular del año 2022, la producción cafetera disminuyó en un 12% con respecto al 2021, debido a las prolongadas lluvias por el fenómeno de la niña, esto repercutió en una disminución de las exportaciones para el mismo año en un 8% (Federación Nacional de Cafeteros de Colombia, 2023). Este comportamiento pone de relieve a su vez que no se tiene una política clara en cuanto a los manejos de inventarios de seguridad para suplir tanto la demanda externa como interna, tal como lo sugiere USDA, (2023a).

Figura 14 Análisis de la producción, exportaciones e inventarios de café verde

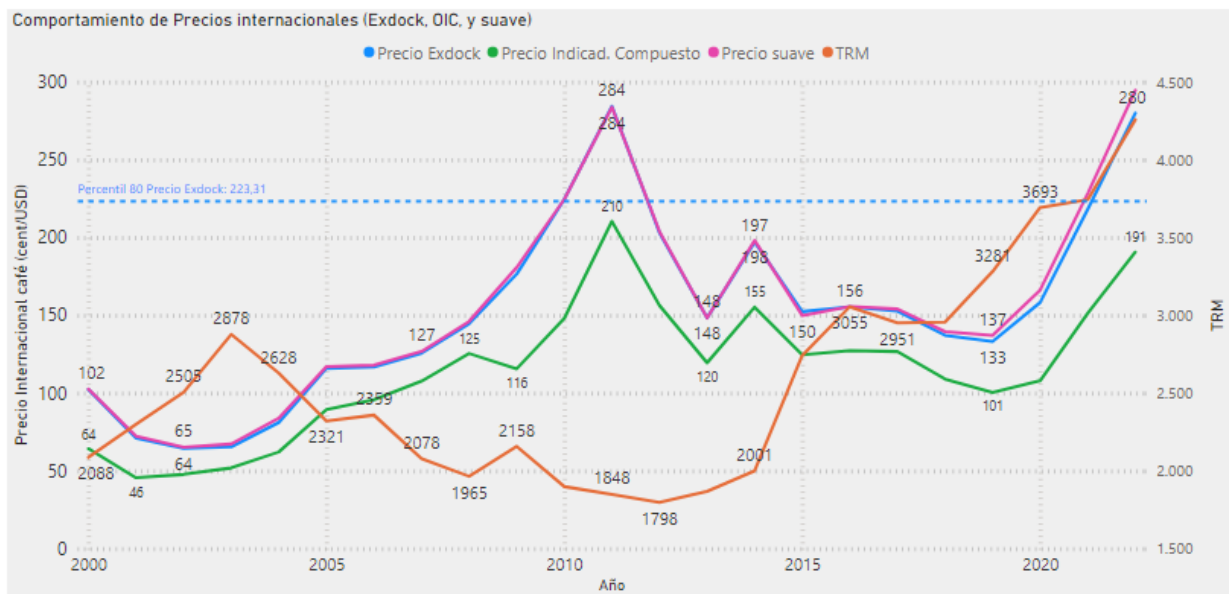


Fuente: Elaboración propia

Por otro lado, revisando más a detalle en la Figura 15, la tendencia de los precios internacionales del café verde dados por el precio Exdock, el precio del indicador compuesto y el precio OIC del café suave colombiano, se puede señalar que presentan una tendencia opuesta al comportamiento de la producción y exportaciones. Lo anterior sugiere que cuando hay mayor producción de café y exportaciones, los precios internacionales tienden a disminuir.

Adicional a lo señalado anteriormente, se evidencia que para el caso del precio Exdock, en los años 2021 y 2022 registra valores por encima del percentil 80⁶ de la serie de datos analizada con valores por encima de los 223,31 centavos US/libra de café verde. Esto puede atribuirse a la escasez de suministro del grano, debido a factores climáticos y a la pandemia del COVID-19, y que conlleva a que todos los actores de la cadena de abastecimiento deban tomar decisiones rápidas que garanticen un abastecimiento fluido y oportuno del café (Pataconi, 2022).

Figura 15 Análisis de precios internacionales del café colombiano



Fuente: Elaboración propia

La base de datos analizada que hace parte del presente estudio tiene 276 registros. En la Tabla 4 mostrada a continuación se observa el resumen descriptivo de las variables que hacen parte este trabajo.

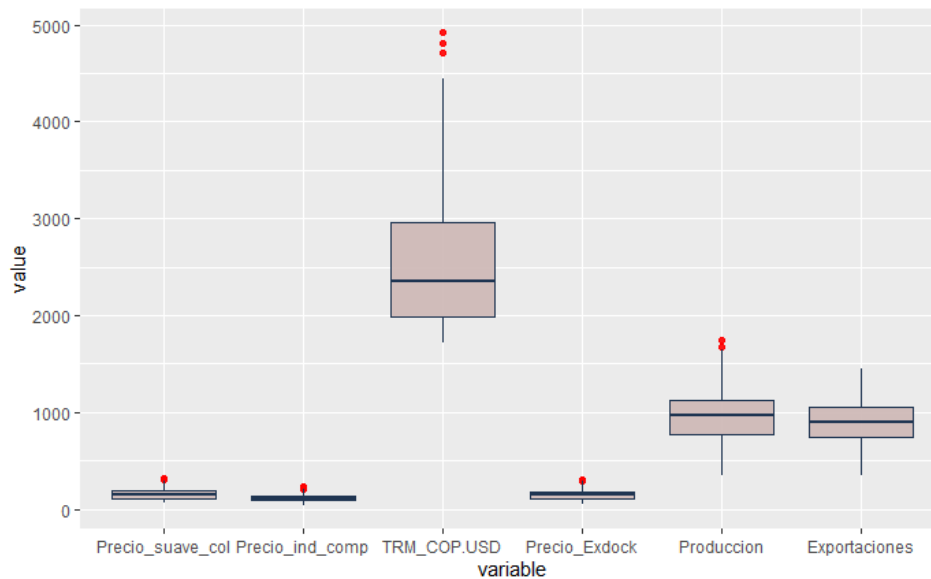
⁶ El percentil 80 significa que el 80% de todos los datos de la serie se encuentran por debajo de dicho valor de referencia. Lo cual, para este caso se traduce en que el precio Exdock en los años 2021 y 2022 se encuentran dentro del 20% más alto de los últimos 22 años.

Tabla 4 Resumen descriptivo de variables

Variable	faltantes	Media	sd	p0	p25	p50	p75	p100	hist
Precio_suave_col	0	154.96	63.53	58.92	113.56	144.95	186.37	319.63	
Precio_ind_comp	0	114.49	44.00	41.17	89.19	113.63	135.46	231.24	
TRM_COP.USD	0	2581.54	685.52	1712.28	1980.72	2351.86	2963.86	4922.30	
Precio_Exdock	0	152.55	62.00	56.92	113.69	143.78	183.70	308.55	
Produccion	0	967.55	261.30	345.00	777.75	975.00	1131.00	1743.00	
Exportaciones	0	894.18	217.25	345.00	745.50	895.50	1057.25	1450.00	

Fuente: Elaboración propia

Figura 16 Diagrama de cajas de variables



Fuente: Elaboración propia

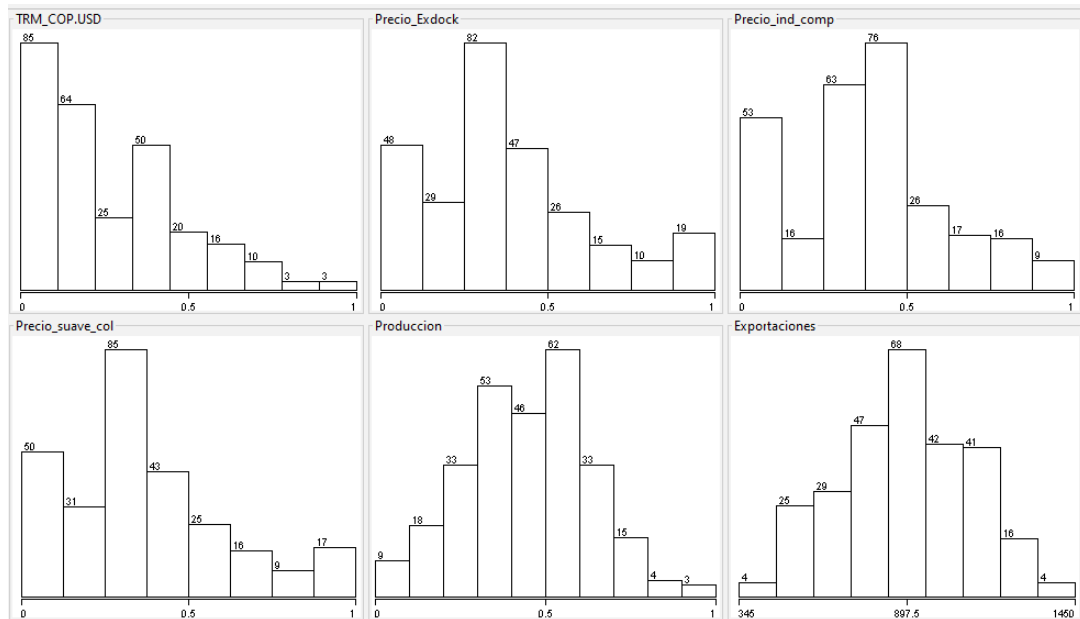
Así mismo, en la Tabla 4 también puede observarse la distribución de las variables seleccionadas. Es así como, variables como la TRM tiene un sesgo a la derecha, es decir,

presenta asimetría positiva. En este caso, la media es mayor a la mediana, por lo que puede indicar que hay valores atípicos altos que estiran la cola hacia la derecha. Este comportamiento también resume lo observado en la variable de precio Exdock, en donde la media es mayor a la mediana, y el sesgo se presenta hacia la derecha (ver Figura 15).

Para la TRM, es evidenciable que la mayoría de los registros en la serie de datos tienen un valor máximo de 3,000 COP/USD, pero se observan valores dispersos por encima del tercer cuartil (percentil 75) que corresponden a valores máximos de repunte de la Tasa Representativa del Mercado. Esto a su vez es observado en la Figura 15, en donde se señala el repunte la TRM indicando una devaluación del peso colombiano respecto al dólar estadounidense (USD), esta tendencia al alza favorece a los productores nacionales independiente de la cantidad exportada debido a que aumenta el precio base del mercado (Valencia y Hurtado, 2021).

Por otro lado, las variables de exportación y producción presentan una distribución casi sin sesgo. Para el caso, de la primera variable, la mediana es muy cercana a la media (ver Figura 17), mientras que, para la segunda variable, hay un sesgo un poco marcado a la izquierda y en donde la media es menor que la mediana.

Figura 17 Histograma de variables analizadas



Fuente: Elaboración propia

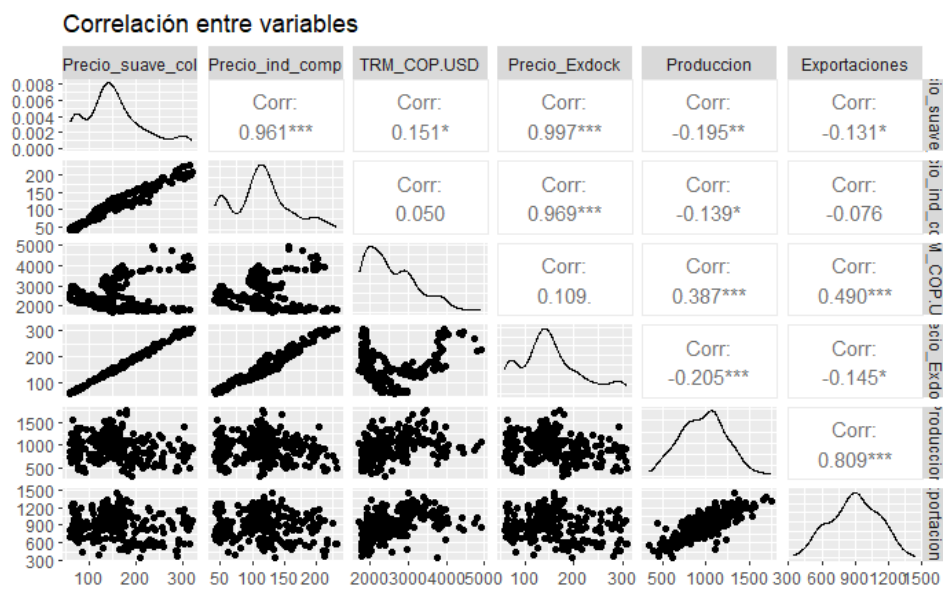
Ahora bien, el análisis de correlación de las variables seleccionadas mostrado en la Figura 18 evidencia que las variables producción y exportación tienen una correlación positiva fuerte, esto indica que a medida que la producción aumenta, la exportación tiende a aumentar. Esta fuerte correlación entre estas dos variables también es evidenciable en el comportamiento de serie de tiempo observado anteriormente en la Figura 14, donde se muestra que la producción y las exportaciones de café verde tienen similar tendencia.

Por su parte, todo lo contrario, es observado con las variables precio Exdock y exportaciones, en donde hay una correlación negativa débil. En términos generales, explica que cuando el precio aumenta, las exportaciones disminuyen, pero como se ilustra en la Figura 18, la relación es débil y no muy significativa. Este mismo comportamiento se presenta entre las variables Precio Exdock y la Producción, la relación es negativa, pero no muy significativa.

Se observa también una correlación entre los precios internacionales del café dados por Precio Exdock, el Precio indicativo compuesto y el Precio del café suave colombiano, esto dado a que son precios internacionales de similares características los cuales son establecidos

por la OIC. Entre dichos precios internacionales existe una correlación negativa, aunque débil con las exportaciones de café verde. Este comportamiento se explica en términos de demanda y oferta de café verde, en ese sentido, cuando aumentan las exportaciones (más oferta) caen los precios internacionales de café verde, mientras que cuando descienden las exportaciones (más demanda), hay un aumento de dichos precios.

Figura 18 Correlación de variables



Fuente: Elaboración propia

Asimismo, en la Figura 18, es notable la existencia de una relación entre las exportaciones de café verde y la tasa de cambio (TRM). Analizando esta correlación a la luz del comportamiento observado en la Figura 14, se evidencia cómo la revaluación del peso frente al dólar provoca una disminución en la cantidad de exportaciones, que, en conjunción con otros factores como las afectaciones producidas por el clima, y la falta de tecnificación de la caficultura colombiana, tienen como resultado una disminución de la participación del grano colombiano en los mercados mundiales.

Considerando los resultados de la caracterización de factores, se destaca la tendencia a la baja en la exportación y producción de café verde colombiano desde 2019, acompañada de un aumento en los precios internacionales, aspecto analizado por Figueroa et al., (2018), este incremento se atribuye, en parte, al alza en los precios de las materias primas, impactando los precios internacionales del grano. Ante esta dinámica, los caficultores adoptan estrategias como la optimización de recursos y la implementación de prácticas mejoradas, ajustándose a estándares de exportación debido al aumento en la demanda (Mehrez et al., 2023).

Para este trabajo se seleccionarán todas las variables ya mencionadas, para lograr una comprensión más profunda y completa de cómo interactúan y afectan la demanda de exportación.

8.3 Preprocesamiento de los datos

El preprocesamiento de los datos implica realizar acciones sobre el conjunto de datos, como estandarizar, normalizar, eliminar datos erróneos o anormales, imputar o eliminar datos faltantes (Contreras-Bravo et al., 2022). Para este trabajo, el conjunto de datos sobre los que se desarrolla la construcción del modelo de ML tiene 276 observaciones sin datos nulos, anormales o erróneos, como se evidenció en el análisis descriptivo.

Por otro lado, se eliminó la variable mes, ya que para efectos de este trabajo no será tenida en cuenta la dependencia temporal en la construcción de los modelos de ML. Asimismo, se realizó estandarización de las covariables (TRM, precio exdock, precio indicativo compuesto, precio suave colombiano y producción), el cual consiste en que los datos tengan media igual a 0 y desviación estándar igual a 1. Esto se realiza debido a que algunos datos pueden añadir más peso que otros y alterar el resultado de algunos algoritmos (Contreras-Bravo et al., 2022). La estandarización no fue realizada sobre la variable de salida (exportaciones). Este proceso es realizado en RStudio mediante la función `scale` de la librería `dp1yr`.

8.4 Construcción del modelo de ML

La construcción del modelo se llevó a cabo siguiendo la estructura propuesta en la Figura 8, que hace parte de la capa de analítica y ML establecido en la arquitectura de inteligencia de negocios planteada en este trabajo. Como se mencionó en la descripción de variables, en esta investigación se cuenta con un conjunto de datos los cuales están conformados por 5 covariables (TRM, precio exdock, precio indicativo compuesto, precio suave colombiano, y producción) y 1 variable respuesta (exportación). Este conjunto de datos previamente ha sido preprocesado, para que sea el insumo para los algoritmos de ML para su entrenamiento y posterior predicción. Cabe mencionar que como herramienta base para la implementación y evaluación de los algoritmos de ML, se utilizó RStudio.

Para el entrenamiento de todos los modelos construidos se seleccionó la validación cruzada repetida, como método para la división de los datos en conjunto de entrenamiento y conjunto de prueba, esto dado sus ventajas, como menor variación de datos y la prevención de sobreajuste o subajuste en el modelo (Contreras-Bravo et al., 2022; Korjus et al., 2016). La implementación de validación cruzada repetida (`method = "repeatedcv"`) se realizó con 10 pliegues (`k-folds = 10`) y 3 repeticiones (`repeats = 3`). Este número de pliegues es el que según la literatura tiene mejor desempeño (Contreras-Bravo et al., 2022; Ghatak, 2017; Korjus et al., 2016).

La evaluación de los algoritmos implementados se realizó basada en métricas como RMSE, R-cuadrado y MAE, estas métricas fueron seleccionadas debido a su gran utilidad en modelos de predicción por regresión en algoritmos de aprendizaje supervisado (Contreras-Bravo et al., 2022; Ghatak, 2017; Korjus et al., 2016; Santana, 2006).

8.4.1 Modelo de regresión lineal múltiple

Se utiliza la biblioteca `caret` para realizar un modelo de regresión lineal múltiple con validación cruzada `k-folds` repetida. Así mismo, se utiliza la función `train` para entrenar un modelo de regresión lineal múltiple (`method = "lm"`) con la columna `EXP` como variable dependiente (variable respuesta) y todas las demás columnas como variables independientes (`EXP ~ .`).

8.4.2 Regresión de Ridge

Para el algoritmo de regresión de Ridge se utilizó el paquete `glmnet` y nuevamente la función `train` de `caret` para realizar validación cruzada. A continuación, se definió la fórmula para el modelo (`ridge_cv <- EXP ~ .`), donde 'EXP' es la variable de respuesta y '.' indica que se deben considerar todas las demás variables como predictores.

El ajuste de los hiperparámetros (`alpha` y `lambda`) se realizó mediante una búsqueda en una cuadrícula (`search = "grid"`), esto implica que se probaron diferentes combinaciones de hiperparámetros para encontrar la configuración óptima. En donde `alpha` se mantiene en 0 (lo que representa Ridge), y `lambda` varía en una secuencia de 0 a 1 con 100 pasos; `lambda` controla la fuerza de la regularización en el modelo Ridge.

8.4.3 Modelo de Red elástica (Elastic Net)

De forma análoga al algoritmo de regresión de Ridge, para el algoritmo de red elástica también se utilizó `glmnet` en conjunto con la función `train` de `caret` para realizar validación cruzada. Subsecuentemente se definió la fórmula del modelo de red elástica (`elastic_cv <- EXP ~ .`), donde `EXP` es la variable de respuesta y el '.' indica que se deben considerar todas las demás variables como predictoras.

A diferencia del algoritmo de regresión de Ridge, la Red elástica utiliza tanto la penalización L1 como la L2. El parámetro `alpha` controla la mezcla entre estas dos penalizaciones. Por

tanto, cuando α es 0, la Red Elástica se comporta como Ridge (usando únicamente penalización L2), y cuando α es 1, se comporta como la regresión LASSO (que utiliza únicamente la penalización L1).

8.4.4 Modelo Máquinas de Soporte Vectorial (SVM)

Para la implementación del algoritmo de SVM se utilizó la librería `e1071` y `caret` en R. Nuevamente se procedió a definir la fórmula del modelo, donde `EXP` es la variable de respuesta y `.'` indicando que se utilizaron todas las demás variables como predictores. Para el entrenamiento del modelo SVM de base radial fue empleada la función `train` del paquete `caret`. Así mismo, se especificó el método `svmRadial`, indicando un SVM con un núcleo radial, lo cual significa que el modelo SVM buscará un modelar relaciones no lineales en los datos.

9 Discusión de resultados

Luego de realizar la implementación de los algoritmos anteriormente mostrados (Regresión lineal múltiple, Regresión de Ridge, Red elástica, y SVM radial) para la construcción de los modelos de predicción de exportación de café, los resultados se exhiben en la Figura 19.

Figura 19 Resultados de modelos

Modelo <chr>	RMSE <dbl>	Rsquared <dbl>	MAE <dbl>
Regresión Lineal	117.6034	0.7176173	91.10323
Ridge	119.2379	0.7103305	94.22443
Elastic Net	117.2899	0.7186866	91.07916
SVM_rad	119.6401	0.6999624	89.33307
4 rows			

Fuente: Elaboración propia

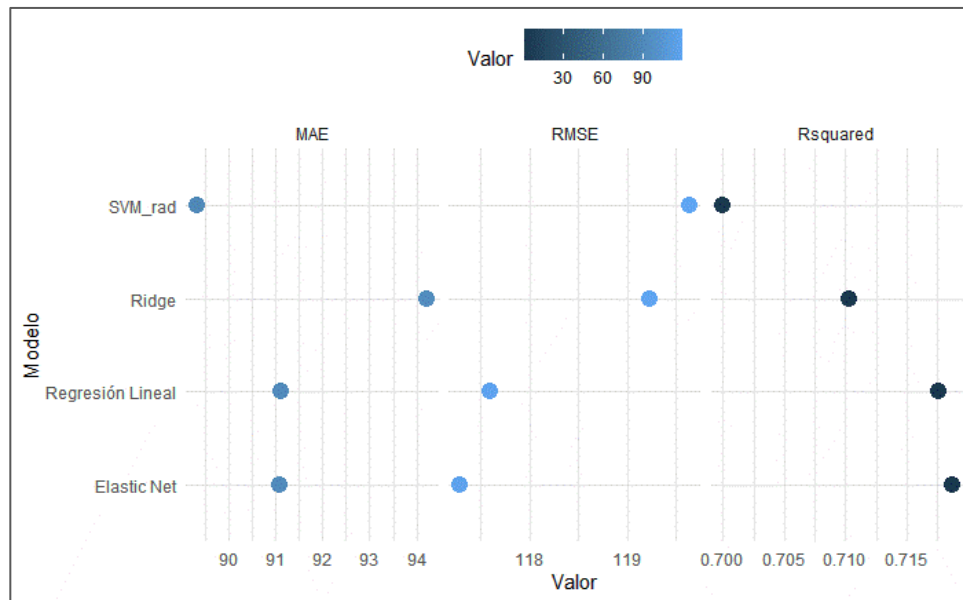
En la Figura 19 es posible observar que los algoritmos de regresión lineal (múltiple, Ridge y Red elástica) así como el modelo de SVM radial presentan valores de RMSE y R-cuadrado similares, lo que indica que son capaces de predecir la demanda de exportación de café con un error relativamente bajo y con un alto grado de precisión. El modelo de Red Elástica destaca con el menor valor de RMSE registrando 117.2899, lo que significa que, en promedio, las predicciones del modelo se desvían alrededor de 117 mil sacos de café de los valores reales para cada observación a lo largo de los años 2000 a 2022. También es evidenciable que los modelos de Regresión lineal múltiple y Red elástica están proporcionando ajustes similares en términos de RMSE, los cuales se encuentran alrededor de los 117.6034 y 117.2899 respectivamente.

Por su parte los valores de R-cuadrado, como se puede observar en la Figura 20, son también bastante similares entre sí. El modelo SVM radial exhibe el menor valor de R-cuadrado, mientras que el modelo de Red elástica obtiene el más alto, alcanzando un 0.71868, lo que quiere decir que las variables predictoras en este modelo explican el 71.868% de la

variabilidad de los datos. A pesar de las diferencias, todos los modelos muestran valores similares de R-cuadrado, alrededor de 0.71, indicando un buen rendimiento y la capacidad de explicar una gran proporción de la variable respuesta.

En cuanto a los valores del MAE, el modelo SVM obtuvo el mejor valor dado por 89.33, esta métrica al igual que el RMSE se expresa en las mismas dimensiones de la variable respuesta, lo que quiere decir, que en términos del MAE para el modelo SVM radial hay una desviación, entre la predicción y valores reales, de 89 mil sacos de café exportados. En el lado opuesto, se encuentra el modelo de Ridge que presentó el valor de MAE más alto de los modelos construidos (ver Figura 20).

Figura 20 Comparación de los resultados

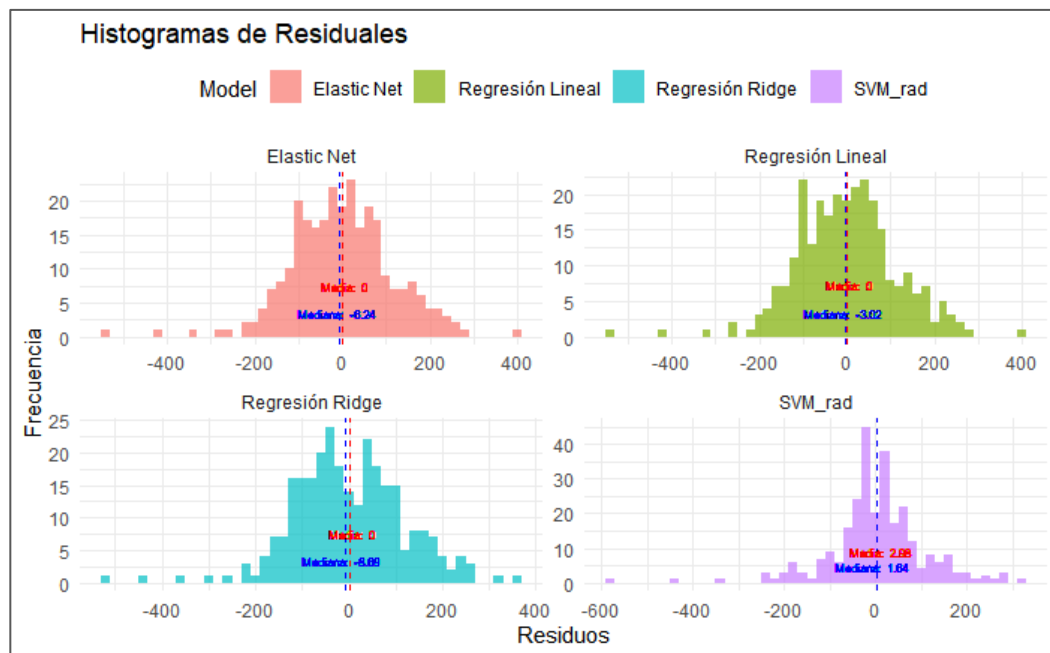


Fuente: Elaboración propia

En cuanto a los residuales obtenidos en cada uno de los modelos construidos, se puede evidenciar en la Figura 21 que los residuos de los modelos de regresión (lineal, Ridge y elástica) no siguen una distribución simétrica alrededor de cero. En este caso, la media es positiva para estos tres modelos lo que implica que se tiende a sobreestimar la variable

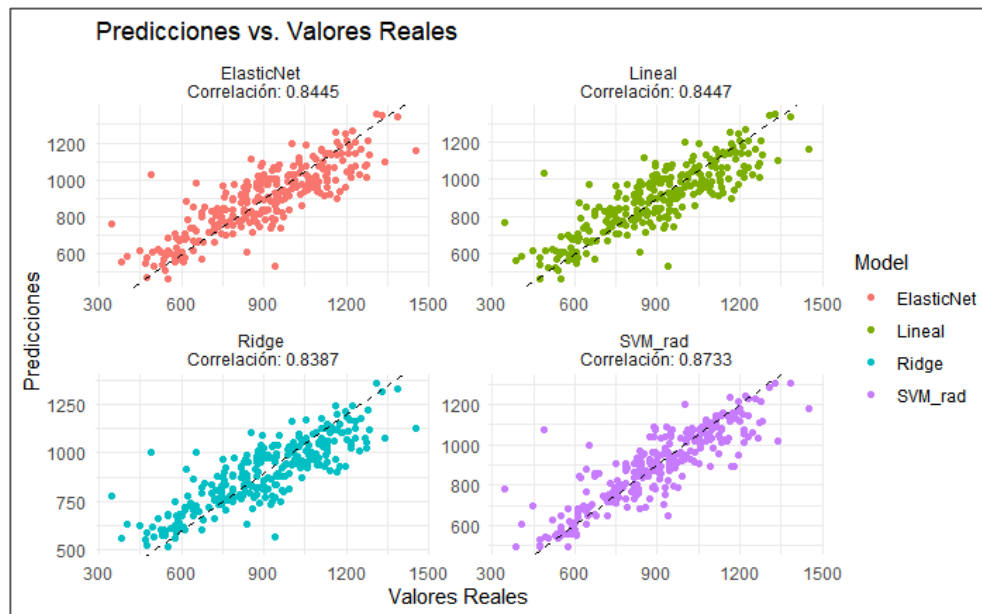
objetivo; mientras que la mediana negativa indica que al menos la mitad de los residuos son negativos, lo que sugiere que hay una cola de residuos más bajos que afecta la mediana (sesgo a la derecha). Por otro lado, en el modelo SVM radial, si bien la media y la mediana no son iguales, se observa una distribución de los residuos que no tiene un sesgo significativo y es relativamente simétrica alrededor de cero.

Figura 21 Histograma de residuales de modelos



Fuente: Elaboración propia

Figura 22 Correlación de predicciones vs. Valores reales de los modelos



Fuente: Elaboración propia

Así mismo, a partir de estos resultados observados en la Figura 22 se puede evidenciar que el modelo SVM registra la mayor correlación entre los modelos evaluados. Aunque este modelo presenta un buen rendimiento, cabe señalar que la diferencia en el rendimiento entre SVM y los modelos de regresión lineal (múltiple, Ridge y Elástica) es mínima. Es decir, los algoritmos de regresión podrían lograr un rendimiento comparable al SVM radial, dependiendo de los datos de entrenamiento y los parámetros del modelo.

Los modelos construidos en este trabajo, según las métricas evaluadas, exhibieron una buena generalización, lo cual los posiciona como herramientas útiles para pronosticar la demanda externa de café en Colombia. Esta cualidad resulta fundamental para su implementación efectiva en entornos reales de predicción de la demanda, especialmente en el caso de las exportaciones de café verde, donde se convierten en un elemento clave para respaldar la toma de decisiones en este sector. En la Figura 23 se ilustra cómo se integra el pronóstico de la demanda al proceso de toma de decisiones, permitiendo la delineación de

diferentes escenarios, tales como aumento, disminución y estabilidad en la demanda, le permitan a la FNC planificar las acciones pertinentes en cada uno de ellos.

Figura 23 Proceso de toma de decisiones para la cadena de suministro del café verde en Colombia

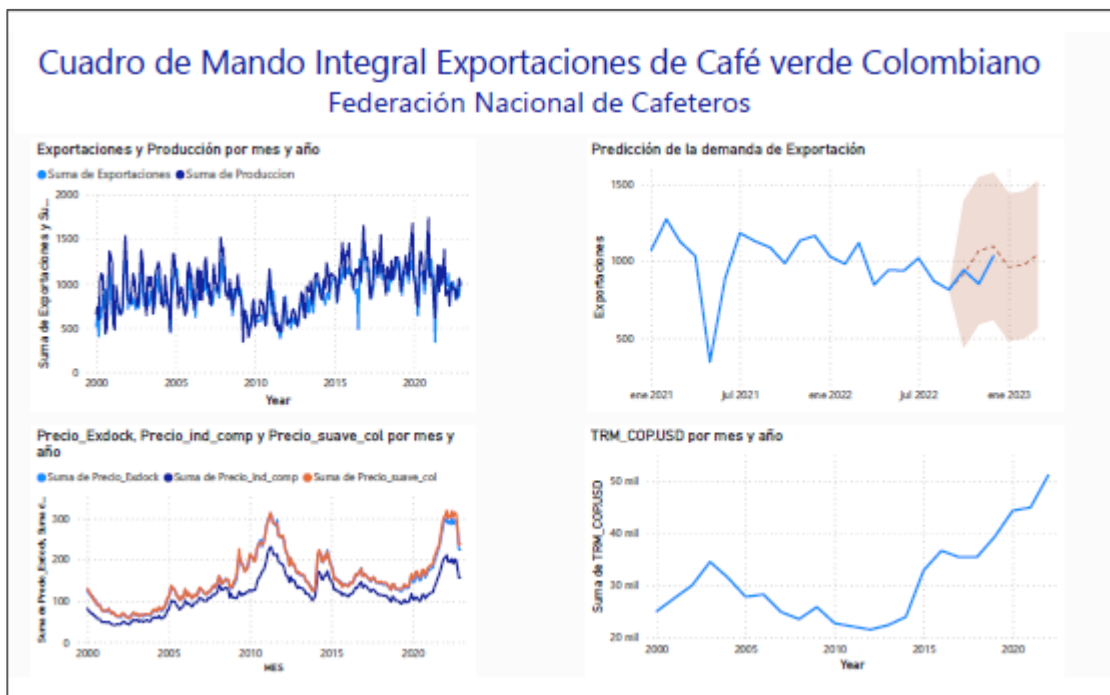


Fuente: Elaboración Propia. Basado en Ensayos sobre economía cafetera - FNC, (2014)

La información resultante del modelo de BI y de predicción de la demanda puede ser aprovechada por la FNC como usuario final. Esta entidad puede acceder a los resultados a través de paneles de control e indicadores clave de desempeño (KPIs por sus siglas en inglés) elaborados en la plataforma Power BI, la cual permite la integración del modelo de predicción desarrollado en RStudio. En la Figura 24 se muestra un ejemplo de dicho tablero que permite monitorear algunas de las variables que influyen en la demanda u otros indicadores de gestión, así como los resultados de las exportaciones reales versus los pronósticos. Estos datos son fundamentales para trazar estrategias óptimas. Este enfoque está alineado con las directrices propuestas por Ramanathan y Ramanathan, (2021); Sharma et al., (2020), quienes destacan

la importancia de implementar estrategias proactivas para mejorar la eficiencia y promover la toma de decisiones informadas en la industria cafetera. Así mismo, los hallazgos obtenidos en este trabajo respaldan los resultados encontrados por Rodríguez et al., (2021), quienes también enfatizan la relevancia de este enfoque para fortalecer el sector cafetero en un contexto dinámico.

Figura 24 Ejemplo de CMI con integrando el modelo de la predicción de la demanda.



Fuente: Elaboración Propia

10 Orientaciones metodológicas para la predicción de la demanda de exportación de café verde en Colombia bajo el enfoque de BI

La arquitectura propuesta de inteligencia de negocio tiene como objetivo principal pronosticar la demanda de exportación de café verde en Colombia y evaluar su impacto en la cadena de suministro. Esta estructura se fundamenta en la inclusión de una capa analítica y la aplicación de cuatro algoritmos de aprendizaje automático. Durante este proceso, se integran variables de entrada como la producción de café verde en Colombia, el precio internacional, el precio OIC, el precio indicativo compuesto y la tasa de cambio, mientras que la variable de respuesta es la cantidad de sacos de café exportados. El propósito de este enfoque es generar información relevante que respalde la toma de decisiones en la cadena de suministro del café. La arquitectura de BI propuesta consta de cinco capas, que se pueden adaptar según las necesidades de los actores de la cadena, esto dependerá del problema a resolver y de los datos disponibles.

Por lo tanto, es esencial que las organizaciones dedicadas a la investigación y apoyo al sector cafetero compartan información de manera colaborativa. Esto permitirá enriquecer el conjunto de datos con más covariables que puedan influir en la predicción de la demanda de exportación. Entre las fuentes de datos sugeridas además de las de datos abiertos de la FNC y el Banco de la República, utilizadas en este trabajo, se sugieren otras fuentes como Cenicafé (Centro de investigación la FNC) y Almacafé como el principal operador logístico de café en Colombia, responsable de recopilar información de gestión de compras, almacenamiento, servicios especializados y aduaneros del grano colombiano (Almacafé, 2018).

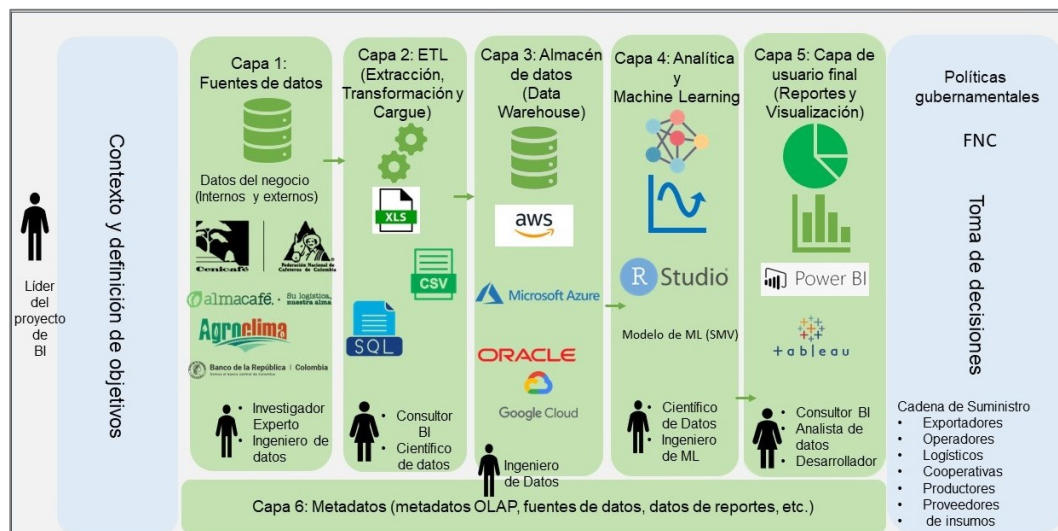
En ese sentido, cobra relevancia la creación de un almacén de datos (DWH, por sus siglas en inglés) gestionado por colaboradores de la FNC o Cenicafé que permita tener un repositorio consolidado de las bases de datos de las fuentes seleccionadas para su posterior análisis. Dentro de los principales proveedores de DWH en el cuadrante mágico de la Consultora

Gartner 2023 se encuentran AWS (Amazon Web Services), Microsoft Azure, Oracle y Google, la elección dependerá de las necesidades específicas de la arquitectura de BI (*Gartner Magic Quadrant & Critical Capabilities* | Gartner, 2023). Una vez recolectados los datos, se procederá con la depuración y el procesamiento mediante ETL, seguido del análisis exploratorio de datos para comprender el comportamiento de las variables.

Posteriormente, se integra en la arquitectura, el modelo de machine learning con el mejor comportamiento para la predicción de la demanda, en este caso, el modelo de Máquina de soporte Vectorial. No obstante, se puede evaluar la utilización de otros algoritmos de predicción como los Redes Neuronales Artificiales de perceptrón multicapa MLP (por sus siglas en inglés), los modelos Gaussianos (GP por sus siglas en inglés) o el refuerzo de gradientes extremos XGBoost, los cuales no fueron evaluados en este estudio, pero que han demostrado tener un buen rendimiento para la predicción en series de tiempo considerando la naturaleza temporal de los datos (Ahmed et al., 2010; Barrera-Animas et al., 2022).

La información derivada de la implementación del modelo predictivo debe integrarse en una plataforma de inteligencia de negocios, como Power BI o Tableau, recomendadas por (Gartner, 2023). Estas plataformas permiten visualizar y compartir de manera efectiva los resultados con los tomadores de decisiones, en este caso la FNC. La Figura 25 presenta una propuesta completa de la arquitectura de BI para la predicción de la demanda en la cadena de suministro del café, contribuyendo a la toma de decisiones para los diversos actores de la cadena y las organizaciones que respaldan el sector cafetero en Colombia.

Figura 25 Arquitectura de BI – propuesta desde las orientaciones metodológicas



Fuente: Elaboración propia

Para lograr una implementación eficaz del modelo de BI en la cadena de suministro del café, se requiere un equipo especializado. Esto incluye un director de proyecto con experiencia relacionada y acceso a diversas fuentes de datos, un consultor de BI para la estructuración del modelo y la integración de sistemas, un científico de datos para la capa de ETL y desarrollo de algoritmos de predicción, apoyado en un Ingeniero de ML, un Ingeniero de datos especializado en la ciberseguridad del almacén de datos, y un desarrollador y un analista de datos que apoyen la capa 5 del modelo de BI. Así mismo, el monitoreo continuo es esencial para evaluar y actualizar el rendimiento del modelo, minimizando los riesgos potenciales.

El éxito de la implementación del modelo de BI depende de la disposición colaborativa de los actores en la cadena de suministro del café, incluyendo productores, cooperativas, plantas de beneficio, comercializadores internos y exportadores. El intercambio colaborativo de información y conocimiento contribuye a mejorar continuamente los procesos, los niveles de producción y la calidad del café verde colombiano. Por último, la transferencia de conocimiento a los participantes de la cadena es esencial para que puedan establecer estrategias que

impulsen el aumento de las exportaciones, desarrollen ventajas competitivas y tomen decisiones informadas sobre la gestión de recursos, aportando así a la sostenibilidad de la cadena del grano.

11 Conclusiones y trabajos futuros

11.1 Conclusiones

Para adelantar esta investigación, se realizó una revisión de la literatura con el objetivo de guiar la elaboración del modelo de predicción de la demanda de exportación de café verde en Colombia desde el enfoque de inteligencia de negocios y cadenas de suministro. El marco teórico abordó temáticas de interés como la industria del café en Colombia, cadena de suministro del café, su sostenibilidad y los factores económicos que influyen en la demanda de exportación del café verde.

Además, se exploraron modelos de inteligencia de negocios y modelos de predicción, consultando bases científicas sobre la implementación de la inteligencia de negocios en el sector cafetero, encontrando 761 publicaciones en los últimos cinco años, profundizando en cinco seleccionados en los cuales se examinaron diferentes enfoques de modelos de predicción, con énfasis en los modelos de machine learning, así como metodologías, variables clave y conclusiones de estudios previos relacionados con la cadena de suministro del café y la inteligencia de negocios. Se encontró que, a pesar de ser un campo en constante crecimiento a nivel internacional, los resultados revelaron una notable escasez de estudios centrados en la demanda de café verde en el contexto colombiano.

Por otro lado, se realizó la caracterización de los factores económicos que inciden en la demanda de exportación del café verde en Colombia lo que ha permitido una comprensión profunda de los determinantes clave que influyen en su comportamiento estadístico. Los factores identificados fueron: la producción, el precio internacional, la tasa de cambio y el volumen de exportación. Se puede evidenciar que existe una fuerte correlación positiva entre los volúmenes de producción y de exportación (con un factor de correlación de 0.809), mostrando una tendencia similar, de lo que se concluye que, a mayor cantidad de sacos de café verde producidos en el país, mayor es la cantidad del grano exportada. También se

comprobó que, cuando aumenta el volumen de café verde exportado (mayor oferta), bajan los precios internacionales, así mismo, cuando descienden las exportaciones, hay un aumento en dichos precios.

La construcción y evaluación de modelos para la predicción de la exportación de café verde en Colombia, utilizando técnicas de aprendizaje automático, destacó la eficacia del modelo SVM, registrando una mayor correlación con respecto a los modelos evaluados. Según las métricas calculadas, este modelo mostró un rendimiento superior con bajos valores de RMSE y MAE, 119.6401 y 89.33307, respectivamente, así como también una notable correlación entre los datos de predicción y valores reales. Aunque este modelo demostró ser eficiente, se observó que la diferencia en el rendimiento con respecto a los modelos de regresión lineal (múltiple, Ridge y Red elástica) es pequeña. Lo anterior sugiere que los algoritmos de regresión lineal pueden ofrecer un rendimiento similar al de SVM, dependiendo de los datos de entrenamiento y de los parámetros del modelo.

La inclusión de modelos de ML, como el SVM, y los otros evaluados en la presente investigación, dentro de una arquitectura de BI, permiten a los diferentes actores de la cadena de suministro prepararse para distintos escenarios. En caso de un aumento en la demanda, los caficultores, cooperativas y otras entidades interesadas pueden ajustar sus estrategias de producción, inventario y exportación, aprovechando oportunidades derivadas de factores como cambios en políticas internacionales, o climáticos, en otras regiones cafeteras. Por otro lado, frente a una contracción en la demanda internacional, es esencial que los participantes en la cadena evalúen cuidadosamente las condiciones del mercado y adopten medidas como la reserva de inventarios para futuras oportunidades de aumento de la demanda o incremento en los precios internacionales.

Finalmente, en este estudio se presentaron las orientaciones metodológicas fundamentales para abordar la predicción de la demanda de exportación de café verde en Colombia desde la

perspectiva de la inteligencia de negocios y cadena de suministro. En este enfoque, se resalta la crucial necesidad de fomentar la colaboración y el intercambio de información entre los actores clave de la cadena. Se subraya, por tanto, la importancia de establecer canales de comunicación fluidos entre productores, cooperativas, entidades como Cenicafé y Almacafé. Así mismo, se presenta una propuesta para complementar la arquitectura de BI de tal manera, que se puedan evaluar nuevas variables, modelos de predicción y herramientas de visualización útiles para los actores que hacen parte de este eslabón de la cadena de suministro del café.

11.2 Trabajos futuros

La presente investigación, enfocada en la evaluación de modelos de predicción para la demanda de exportación de café verde en Colombia, incorporó factores económicos como la producción, los precios internacionales y la TRM como covariables. Sin embargo, existen oportunidades para futuros estudios que amplíen el análisis mediante la inclusión de otras covariables significativas, tales como factores climáticos, áreas cultivadas o la calidad del grano.

En ese sentido, se invita en próximos trabajos a explorar la implementación de otros algoritmos de machine learning para regresión, como Redes Neuronales o Bosque Aleatorio, involucrando series de tiempo ya que el factor de la temporalidad no se abordó en este estudio, lo que incluiría un ajuste en el modelo de almacenamiento de datos, para que este sea orientado a un mayor volumen de datos temporales (Ahmed et al., 2010; Barrera-Animas et al., 2022; Khatibi et al., 2020).

El conocimiento prospectivo sobre la cantidad de sacos que serán exportados, al emplear diferentes algoritmos de ML junto con el pronóstico climático en las fincas productoras de café,

no solo en Colombia sino también en competidores directos como Brasil, facilitaría la evaluación y la adopción de medidas estratégicas por parte de los actores de la cadena de suministro del café colombiano. Estos actores incluyen caficultores, cooperativas, operadores logísticos (como Almacafé) y exportadores.

Asimismo, se destaca la importancia de enriquecer la arquitectura propuesta en este estudio mediante la creación de herramientas de visualización beneficiosas para diversos actores, como comercializadores, exportadores y la FNC. Este proceso debe estar respaldado por prácticas sólidas de gestión del cambio y la transmisión de conocimientos, facilitando así el respaldo a la toma de decisiones en esta cadena agroindustrial.

Finalmente, otra línea de investigación futura que se deriva de este estudio podría involucrar la aplicación de modelos de inteligencia de negocios y predicción de la demanda en sectores exportadores (tanto tradicionales como no tradicionales), dentro de la economía colombiana. En estos sectores, donde inciden factores económicos similares, la implementación de un modelo de inteligencia empresarial podría proporcionar una valiosa herramienta para la toma de decisiones informadas.

12 Referencias

- Abu-Alsondos, I. A. (2023). The impact of business intelligence system (BIS) on quality of strategic decision-making. *International Journal of Data and Network Science*, 7(4), 1901–1912. <https://doi.org/10.5267/j.ijdns.2023.7.003>
- Acevedo Amorocho, A., Ramírez Carreño, F., Salcedo Blanco, D., & Román Ordóñez, J. (2020). Pronóstico del precio del café : Una propuesta desde los modelos econométricos . *Revista Venezolana de Gerencia*, 25(4). <https://www.produccioncientificaluz.org/index.php/rvg/article/view/35210>
- Ahmed, N. K., Atiya, A. F., El Gayar, N., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5), 594–621. <https://doi.org/10.1080/07474938.2010.481556>
- Almacafé. (2018). *Almacafé. Su logística nuestra alma. Quiénes somos.* <https://www.almacafe.com.co/nosotros/#:~:text=Somos,8%20de%20mayo%20de%201965>.
- Azevedo, A., & Santos, M. F. (2009). *BUSINESS INTELLIGENCE State of the Art, Trends, and Open Issues.*
- Bacci, L. A., Mello, L. G., Incerti, T., Paulo de Paiva, A., & Balestrassi, P. P. (2019). Optimization of combined time series methods to forecast the demand for coffee in Brazil: A new approach using Normal Boundary Intersection coupled with mixture designs of experiments and rotated factor scores. *International Journal of Production Economics*, 212(March 2018), 186–211. <https://doi.org/10.1016/j.ijpe.2019.03.001>
- Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, 7, 100204. <https://doi.org/10.1016/j.mlwa.2021.100204>

- Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. En *Sensors* (Vol. 21, Número 11). MDPI AG. <https://doi.org/10.3390/s21113758>
- Borrella, I., Mataix, C., & Carrasco-Gallego, R. (2015). Smallholder Farmers in the Speciality Coffee Industry: Opportunities, Constraints and the Businesses that are Making it Possible. *IDS Bulletin*, 46(3), 29–44. <https://doi.org/10.1111/1759-5436.12142>
- Cafés Suaves - Federación Nacional de Cafeteros*. (2019). Federación Nacional de Cafeteros. <https://federaciondecafeteros.org/wp/glosario/cafes-suaves/>
- Centro Nacional de Investigaciones de Café. (2021). Guía más agronomía, más productividad, más calidad. En *Cenicafé* (3a ed., Vol. 6, Número August). <https://doi.org/https://doi.org/10.38141/cenbook-0014>
- Chopra, S., & Meindl, P. (2013). Administración de la cadena de suministro. En *Estrategia, planeación y operación*.
- Coffee Report and Outlook*. (2023). https://icocoffee.org/documents/cy2022-23/Coffee_Report_and_Outlook_April_2023_-_ICO.pdf
- Contreras-Bravo, L. E., Nieves-Pimiento, N., & Gonzalez-Guerrero, K. (2022). Prediction of University-Level Academic Performance through Machine Learning Mechanisms and Supervised Methods. *Ingeniería*, 28(1), e19514. <https://doi.org/10.14483/23448393.19514>
- Cuevas, V., Alvares, S., Azcona, M., & Rodríguez, I. (2019). Predictive power of the Support Vector Machine. An application to the financial planning. *Revista Cubana de Ciencias Informáticas*, 13(3), 59–75. <http://rcci.uci.cu>
- Deina, C., do Amaral Prates, M. H., Alves, C. H. R., Martins, M. S. R., Trojan, F., Stevan, S. L., & Siqueira, H. V. (2022). A methodology for coffee price forecasting based on extreme

learning machines. *Information Processing in Agriculture*, 9(4), 556–565.

<https://doi.org/10.1016/j.inpa.2021.07.003>

División de Estadística de la ONU. (2022). *UN Comtrade Analytics*.

<https://comtradeplus.un.org/Visualization/Labs>

Ereth, J., & Baars, H. (2011). A CAPABILITY APPROACH FOR DESIGNING BUSINESS INTELLIGENCE AND ANALYTICS ARCHITECTURES. *Hawaii International Conference on System Sciences*, 53, 5349–5358.

<http://www.ibimapublishing.com/journals/CIBIMA/cibima.html>

Estadísticas cafeteras - Federación Nacional de Cafeteros de Colombia. (s/f). Federación Nacional de Cafeteros. <https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>

Federación Nacional de Cafeteros. (2014). Ensayos sobre Economía Cafetera. *Ensayos sobre economía cafetera*, 30, 1–125.

https://federaciondecafeteros.org/app/uploads/2019/12/Economía-Cafetera-No.-29_Web.pdf

Federación Nacional de Cafeteros de Colombia. (s/f). *Estadísticas cafeteras*. <https://federaciondecafeteros.org/wp/estadisticas-cafeteras/>

Federación Nacional de Cafeteros de Colombia. (2022). *Diversidad del café de Colombia*. <https://federaciondecafeteros.org/wp/productos/cafe-verde/>

Federación Nacional de Cafeteros de Colombia. (2023). *Producción anual de café de Colombia cierra 2022 en 11,1 millones de sacos*.

[https://federaciondecafeteros.org/wp/listado-noticias/produccion-anual-de-cafe-de-colombia-cierra-2022-en-111-millones-de-](https://federaciondecafeteros.org/wp/listado-noticias/produccion-anual-de-cafe-de-colombia-cierra-2022-en-111-millones-de-sacos/#:~:text=Bogot%C3%A1%2C%20enero%20%20de%202023,millones%20de%20sacos%20de%202021)

[sacos/#:~:text=Bogot%C3%A1%2C%20enero%20%20de%202023,millones%20de%20sacos%20de%202021](https://federaciondecafeteros.org/wp/listado-noticias/produccion-anual-de-cafe-de-colombia-cierra-2022-en-111-millones-de-sacos/#:~:text=Bogot%C3%A1%2C%20enero%20%20de%202023,millones%20de%20sacos%20de%202021).

- Figueroa, E., Pérez, F., Godínez, L., & Pérez, R. (2018). Los precios de café en la producción y las exportaciones a nivel mundial. *Revista Mexicana de Economía y Finanzas Nueva Época*. <https://doi.org/https://doi.org/10.21919/remef.v14i1.358>
- Gartner Magic Quadrant & Critical Capabilities | Gartner. (2023). Gartner. <https://www.gartner.com/en/research/magic-quadrant>
- Ghatak, A. (2017). Machine Learning with R. En Springer (Ed.), *Machine Learning with R*. <https://doi.org/10.1007/978-981-10-6808-9>
- Gil, J. G. (2012). Estimación de un pronóstico de exportaciones de café suave colombiano: redes neuronales artificiales y ardl. *Universidad EAFIT*. https://repository.eafit.edu.co/bitstream/handle/10784/11485/JuanGabriel_GilSerna_2016.pdf?sequence=2
- González, R., Barrientos, A., Toapanta, M., & Del Cerro, J. (2017). Aplicación de las Máquinas de Soporte Vectorial (SVM) al diagnóstico clínico de la Enfermedad de Párkinson y el Temblor Esencial. *RIAI - Revista Iberoamericana de Automatica e Informatica Industrial*, 14(4), 394–405. <https://doi.org/10.1016/j.riai.2017.07.005>
- Grupo de Estudios Económicos-Superintendencia de Industria y Comercio. (2012). *Estudios Económicos Sectoriales*. 5. https://www.sic.gov.co/recursos_user/documentos/Estudios-Academicos/Documentos-Elaborados-Grupo-Estudios-Economicos/5_Estudio_Sobre_Sector_Cafe_Colombia_Diciembre_2012.pdf
- Hans, C. (2011). Elastic net regression modeling with the orthant normal prior. *Journal of the American Statistical Association*, 106(496), 1383–1393. <https://doi.org/10.1198/jasa.2011.tm09241>
- Heizer, J., & Render, B. (2012). Dirección de la producción y de operaciones. En *Foreign Affairs* (Vol. 91, Número 5).

Heizer, J., Render, B., Murrieta, J., & Haaz, G. (2009). *Principios de administración de operaciones* (7a ed.). Pearson Educación.

History of Robusta. (2013). World Coffee Research. <https://varieties.worldcoffeeresearch.org/robusta-2/history-of-robusta>

Hofmann, E., & Rutschmann, E. (2018). Big data analytics and demand forecasting in supply chains: a conceptual analysis. *International Journal of Logistics Management*, 29(2), 739–766. <https://doi.org/10.1108/IJLM-04-2017-0088>

International Labour Organization. (2022). *Decent work challenges and opportunities in the coffee supply chain in Colombia*. www.ifrro.org

Jiménez-Varón, C., & Tabares-Soto, R. (2018). Dinámica no lineal del mercado del café en Colombia Non-linear Dynamics of the Coffee Market in Colombia. *Revista Mutis*, 8(2), 73–83. <https://doi.org/10.21789/22561498.1405>

Joyanes, L. (2019). *Inteligencia de negocios y analítica de datos Una visión global de Business Intelligence & Analytics* (Alfa Omega, Ed.).

Khatibi, V., Keramati, A., & Shirazi, F. (2020). Deployment of a business intelligence model to evaluate Iranian national higher education. *Social Sciences and Humanities Open*, 2(1), 100056. <https://doi.org/10.1016/j.ssaho.2020.100056>

Kittichotsatsawat, Y., Jangkrajarn, V., & Tippayawong, K. Y. (2021). Enhancing coffee supply chain towards sustainable growth with big data and modern agricultural technologies. *Sustainability (Switzerland)*, 13(8), 1–20. <https://doi.org/10.3390/su13084593>

Kittichotsatsawat, Y., Tippayawong, N., & Tippayawong, K. Y. (2022). Prediction of arabica coffee production using artificial neural network and multiple linear regression techniques. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-18635-5>

- Korjus, K., Hebart, M. N., & Vicente, R. (2016). An Efficient Data Partitioning to Improve Classification Performance While Keeping Parameters Interpretable. *PLoS One*, 11(8), e0161788. <https://doi.org/10.1371/journal.pone.0161788>
- Krajewski, L. J., Malhotra, M. K., & Ritzman, L. P. (2013). *Administración de operaciones: procesos y cadenas de suministros*. Pearson Educación.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.
- López, G. A., Correa, M. J., & Estrada-Márquez, M. L. (2021). Caracterización de la cadena de suministro de los cafés especiales de Belén de Umbría, Risaralda, Colombia. *Universidad Tecnológica de Pereira*. <https://doi.org/https://doi.org/10.22517/23447214.23911>
- Martínez Chimbi, W. (2019). El café, una auténtica apuesta por sobrevivir. Análisis de las transformaciones económicas, sociales y laborales del sector cafetero en el Viejo Caldas, una región como patente expresión de la crisis. *Universidad Externado de Colombia*, 1.
- Mehrez, K., Khemira, H., & Medabesh, A. (2023). Marketing strategies for value chain development: Case of Khawlani coffee-Jazan Region, Saudi Arabia. *Journal of the Saudi Society of Agricultural Sciences*, 22(7), 449–460. <https://doi.org/10.1016/J.JSSAS.2023.04.004>
- Mejía, C., & Rincón, J. (2021). *Comportamiento de la producción de café en Colombia usando técnicas de analítica de datos*.
- Mejía, J. (2023). *Fundamentos de cadena de suministro: Teorías y aplicaciones* (Primera ed). Academia Mexicana de Investigación y Docencia en Investigación.
- Monsalve, D. (2016). El café colombiano y su exportación. En *Colombia cafetera*.
- Montoya, E., Arcila, J., Jaramillo, A., Riaño, N., & Quiroga, F. (2009). *Modelo para simular la producción potencial del cultivo del café en Colombia*.

- Moreno, E., & Cristiano, G. (2021). *Técnicas de Regularización en el marco del Aprendizaje de Máquina: Regresiones Ridge y Lasso*.
<https://repository.udistrital.edu.co/bitstream/handle/11349/28658/Morenoquinteroemanuellealejandrocristianomu%C3%B1ozgisellnatalia2021.pdf?sequence=2>
- Muñoz-Belalcázar, J. A., Benavides-Cardona, C. A., Lagos-Burbano, T. C., & Criollo-Velázquez, C. P. (2021). Agronomic management on the yield and quality of coffee (*Coffea arabica*) Castillo variety in Nariño, Colombia. *Agronomía Mesoamericana*, 32(3), 750–763. <https://doi.org/10.15517/AM.V32I3.44403>
- Ong, I., Siew, P., & Wong, S. (2011). A Five-Layered Business Intelligence Architecture. *Communications of the IBIMA*, 2011, 1–11. <https://doi.org/10.5171/2011.695619>
- Oviedo-Celis, R. A., & Castro-Escobar, E. S. (2021). Un análisis comparativo de la sostenibilidad de sistemas para la producción de café en fincas de Santander y Caldas, Colombia. *Ciencia & Tecnología Agropecuaria*, 22(3), e2330. https://doi.org/10.21930/rcta.vol22_num3_art:2230
- Patacconi, G. (2022). *Sourcing diversification*.
- Porter, M. E. (2007). *ESTRATEGIA COMPETITIVA Técnicas para el análisis de los sectores industriales y de la competencia*.
- POSCOSECHA - *Café de Colombia*. (2020a). Federación Nacional de Cafeteros. <https://www.cafedecolombia.com/particulares/poscosecha/>
- POSCOSECHA - *Café de Colombia*. (2020b). Federación Nacional de Cafeteros. <https://www.cafedecolombia.com/particulares/poscosecha/>
- Rainforest Alliance. (2021). *UTZ Certification (Now part of the Rainforest Alliance) | Rainforest Alliance*. <https://www.rainforest-alliance.org/utz/>
- Ramanathan, U., & Ramanathan, R. (2021). Information Sharing and Business Analytics in Global Supply Chains. En R. Vickerman (Ed.), *International Encyclopedia of*

- Transportation* (pp. 71–75). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-08-102671-7.10222-2>
- Rayed, C. A. (2019). Using business intelligence solutions for forecasting in marketing researches. *International Journal of Informatics and Communication Technology (IJ-ICT)*, 8(2), 102. <https://doi.org/10.11591/ijict.v8i2.pp102-110>
- Regiones cafeteras - Café de Colombia*. (2020). Café de Colombia. <https://www.cafedecolombia.com/particulares/regiones-cafeteras/>
- Rodríguez, J. P., Montoya-Munoz, A. I., Rodriguez-Pabon, C., Hoyos, J., & Corrales, J. C. (2021). IoT-Agro: A smart farming system to Colombian coffee farms. *Computers and Electronics in Agriculture*, 190. <https://doi.org/10.1016/j.compag.2021.106442>
- Sánchez, N. (2016). Máquinas de soporte vectorial y redes neuronales artificiales en la predicción del movimiento USD/COP spot intradiario. *ODEON*, 9, 113. <https://doi.org/10.18601/17941113.n9.04>
- Sanders, N. R., & Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega*, 31(6), 511–522. <https://doi.org/10.1016/j.omega.2003.08.007>
- Santana, J. C. (2006). Predicción de series temporales con redes neuronales: Una aplicación a la inflación Colombiana. *Revista Colombiana de Estadística*, 29(1), 77–92.
- Shafii, Y., & Ahmad, R. (2014). *Information and Knowledge Management Creating Business Intelligence through Machine Learning: An Effective Business Decision Making Tool*. 4(1). www.iiste.org
- Sharma, R., Kamble, S. S., Gunasekaran, A., Kumar, V., & Kumar, A. (2020). A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Computers and Operations Research*, 119, 104926. <https://doi.org/10.1016/j.cor.2020.104926>

- Taha, H. (2017). *Investigación de operaciones* (10a ed.). Pearson educación.
- United States Department of Agriculture. (2023a). *Coffee Annual Production - Colombia*.
- United States Department of Agriculture. (2023b). *Foreign Agricultural Service/USDA*.
<https://apps.fas.usda.gov/psdonline/app/index.html#/app/downloads>
- Valencia, M. F., & Hurtado, I. H. (2021). *Precio internacional del café, cambio climático y mercados financieros*.
- Velasco, N. M., Villa Betancur, S., & Camacho Ahumada, S. M. (2023). *Gestión de la cadena de abastecimiento: eslabones, herramientas y tendencias*. Universidad de los Andes.
<https://elibro.net/es/lc/bibliotecaean/titulos/246219>
- Velasco Rodríguez, N. M., Villa Betancur, S., & Camacho Ahumada, S. M. (2023). *Gestión de la cadena de abastecimiento: eslabones, herramientas y tendencias*. Universidad de los Andes. <https://elibro.net/es/lc/bibliotecaean/titulos/246219>
- Villarreal, D. F. (2016). "Introducción a los Modelos de Pronósticos".
- Von Lücken, J. (2021). *Métodos de Regularización Lasso, Ridge y Elastic Net: Una aplicación a los seguros de no vida*.
- Wang, C. N., Yu, M. C., Ho, N. N. Y., & Le, T. N. (2021). An integrated forecasting model for the coffee bean supply chain. *Applied Economics*, 53(28), 3321–3333.
<https://doi.org/10.1080/00036846.2021.1887447>
- Watson, H. J. (2009). Business intelligence: Past, present, and future. *Communications of the Association for Information Systems*, 25(1), 487–510.
<https://doi.org/10.17705/1cais.02539>
- World Bank. (2016). Overview of the Global Coffee Sector Supply Chain. *World Bank's Agriculture Global Practice Discussion Paper*, 1–31.

13 Anexos

A continuación, se presenta como anexo el HTML de la construcción y evaluación de los modelos realizada en RStudio.

https://universidadeaneducos.sharepoint.com/:u:/s/504TG_PadillaAgamezBeltran/EdK26gL92fVHi9MxDIdRG40BO04u17LjRcllg8dfd3ycAQ?e=gpnkbr